

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

## SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: Bakur Examiner #: 76550 Date: 10/28  
 An Unit: 1645 Phone Number 30 8-8886 Serial Number: 091868,987  
 Mail Box and Bldg/Room Location: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*  
 Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc. if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Filing Date: \_\_\_\_\_

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Search seq 14  
 against nucleic acid ds  
 except pending

Ph Ar Sy D3  
 Socher cluck 13  
 Road T of Lon 20  
 Vanilla Collection

1 20  
 3h 3h  
 4  
 5p  
 6p  
 6h

## STAFF USE ONLY

| Searcher                                | Type of Search           | Vendors and cost where applicable |
|---|--------------------------|-----------------------------------|
| <u>D. Schneider</u>                     | NA Sequence (#) <u>1</u> | STN _____                         |
| Searcher Phone # <u>308-4292</u>        | AA Sequence (#) _____    | Dialog _____                      |
| Searcher Location <u>CM 6403</u>        | Structure (#) _____      | Questel/Orbit _____               |
| Date Searcher Picked Up <u>10/29/03</u> | Bibliographic _____      | Dr. Linn _____                    |
| Date Completed <u>10/29/03</u>          | Litigation _____         | Lexis Nexis _____                 |
| Searcher Prep & Review Time <u>12</u>   | Fulltext _____           | Sequence Systems <u>CompuGen</u>  |
| Client Prep Time _____                  | Patent Family _____      | WWW Internet _____                |
| Online Time <u>6</u>                    | Other _____              | Other (specify) _____             |

BEST AVAILABLE COPY

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 28, 2003, 15:26:12 ; Search time 94 Seconds  
(without alignments)  
2591.952 Million cell updates/sec

US-09-868-987-14

Perfect score: 2778

Sequence: 1 MVSSPILNVLKPNHASVSK.....SLYIAPPLLEFMRKNSK 552

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spo3/US09868987/runat 28102003\_152602\_23852/app query.fasta\_1.711  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09868987@cgn 1.1.56 @runat 28102003\_152602\_23852 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT\_DSBLCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq:  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq:  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq:  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq:  
5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq:  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length  | ID | Description         |
|------------|-------|-------------|---------|----|---------------------|
| C 1        | 2730  | 98.3        | 1230025 | 4  | US-09-198-452A-1    |
| C 2        | 719   | 25.9        | 1830121 | 4  | US-09-557-884-1     |
| C 3        | 719   | 25.9        | 1830121 | 4  | US-09-643-990A-1    |
| C 4        | 678.5 | 24.4        | 2211    | 3  | US-09-462-844-1     |
| C 5        | 672   | 24.2        | 6321    | 4  | US-09-221-017B-311  |
| C 6        | 646.5 | 23.3        | 2301    | 4  | US-09-134-001C-852  |
| C 7        | 512   | 18.4        | 4403765 | 3  | US-09-103-840A-2    |
| C 8        | 509   | 18.3        | 4411529 | 3  | US-09-103-840A-1    |
| C 9        | 483.5 | 17.4        | 40429   | 4  | US-08-311-731A-125  |
| C 10       | 425   | 15.3        | 1398    | 4  | US-09-252-991A-6413 |
| C 11       | 418   | 15.0        | 1947    | 4  | US-09-328-352-1027  |
| C 12       | 417   | 15.0        | 1890    | 4  | US-09-252-991A-6330 |

FILE COPY

FILE COPY

ALIGNMENTS

RESULT 1

Sequence 1, Application US/09198452A  
Patent No. 6559294  
GENERAL INFORMATION:  
APPLICANT: Griffiths, R.  
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection  
FILE REFERENCE: 9710-003-999  
CURRENT APPLICATION NUMBER: US/09/198.452A  
CURRENT FILING DATE: 1998-11-24  
NUMBER OF SEQ ID NOS: 6849  
SEQ ID NO 1  
LENGTH: 1230025  
TYPE: DNA  
ORGANISM: Chlamydia pneumoniae  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)..(15000)  
NAME/KEY: misc feature  
LOCATION: (15001)..(30000)  
OTHER INFORMATION: n= a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (30001)..(45000)  
OTHER INFORMATION: n= a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (45001)..(60000)  
OTHER INFORMATION: n= a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (60001)..(75000)  
OTHER INFORMATION: n= a or c or g or t  
NAME/KEY: misc feature  
LOCATION: (75001)..(90000)  
OTHER INFORMATION: n= a or c or g or t

Sequence 6245, Ap  
Sequence 6054, Ap  
Sequence 6217, Ap  
Sequence 6138, Ap  
Sequence 1243, Ap  
Sequence 1, Appli  
Sequence 4040, Ap  
Sequence 4006, Ap  
Sequence 3906, Ap  
Sequence 1348, Ap  
Sequence 897, App  
Sequence 2168, Ap  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 14559, A  
Sequence 14791, A  
Sequence 829, App  
Sequence 191, App  
Sequence 14790, A  
Sequence 14540, A  
Sequence 14823, A  
Sequence 443, App  
Sequence 1012, Ap  
Sequence 418, App  
Sequence 10659, A  
Sequence 10414, A  
Sequence 10504, A  
Sequence 14948, A  
Sequence 14688, A  
Sequence 15066, A  
Sequence 319, App  
Sequence 102, App

13 369.5 13.3 969 4 US-09-252-991A-6245  
C 14 369.5 13.3 984 4 US-09-252-991A-6054  
C 15 358 12.9 429 4 US-09-252-991A-6217  
C 16 314 11.3 870 4 US-09-252-991A-6138  
C 17 306.5 11.0 996 4 US-09-328-352-1243  
C 18 170.5 6.1 1664976 4 US-08-916-421B-1  
C 19 138 5.0 3102 4 US-09-252-991A-4040  
C 20 138 5.0 4284 4 US-09-252-991A-4006  
C 21 138 5.0 4704 4 US-09-252-991A-3906  
C 22 137.5 4.9 3120 4 US-09-328-352-1348  
C 23 136 4.9 3096 5 PCT-US96-05320A-897  
C 24 136 4.9 3156 4 US-09-134-001C-2168  
C 25 136 4.9 1830121 4 US-09-557-884-1  
C 26 136 4.9 1830121 4 US-09-643-990A-1  
C 27 132 4.8 1664976 4 US-08-916-421B-1  
C 28 128.5 4.6 3201 4 US-09-252-991A-14559  
C 29 128.5 4.6 3402 4 US-09-252-991A-14791  
C 30 127 4.6 288 4 US-09-134-001C-829  
C 31 126.5 4.6 15393 4 US-09-453-702B-191  
C 32 124 4.5 3003 4 US-09-252-991A-14790  
C 33 124 4.5 3132 4 US-09-252-991A-14540  
C 34 124 4.5 3411 4 US-09-252-991A-14823  
C 35 123.5 4.4 2862 4 US-09-252-991A-443  
C 36 123.5 4.4 3102 4 US-09-328-352-1012  
C 37 123.5 4.4 3156 4 US-09-252-991A-418  
C 38 123 4.4 2862 4 US-09-252-991A-10659  
C 39 123 4.4 2910 4 US-09-252-991A-10414  
C 40 123 4.4 3111 4 US-09-252-991A-10504  
C 41 122.5 4.4 1713 4 US-09-252-991A-14948  
C 42 122.5 4.4 3144 4 US-09-252-991A-14688  
C 43 122.5 4.4 3198 4 US-09-252-991A-15066  
C 44 121.5 4.4 3186 4 US-09-328-352-319  
C 45 118.5 4.3 7886 4 US-09-453-702B-102



```

; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (825001)..(840000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (840001)..(855000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (855001)..(870000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (870001)..(885000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (885001)..(900000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (900001)..(915000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature

```

## Alignment Scores:

```

Pred. No.: 0 Length: 1230025
Score: 2730.00 Matches: 551
Percent Similarity: 99.28% Conservative: 0
Best Local Similarity: 99.28% Mismatches: 1
Query Match: 98.27% Indels: 3
DB: 4 Gaps: 0

```

US-09-868-987-14 (1-552) x US-09-198-452A-1 (1-1230025)

```

Qy 1 MetValSerSerProfileLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20
Db 661906 ATGGTCAGAGCCCTATTATTAACCGTCCCATTTGAAATAATCATGCCAGTGCTCAGGGAAA 661847
Qy 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40
Db 661846 TTTACCCACCGTGAAGTAGGACAACTCGCCTCAGATTTTAAATCTGGAGCGATGCTTTT 661787
Qy 41 ValProGluValLeuSerGluThrIleSerSerAspLeuGlyLysGlnCysThr 60
Db 661786 GTTCCCGAGGTTCTCAGTGAAGAGACGATCTCTCTGATCTTTGGGAAAAAACAATGTACA 661727
Qy 61 GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80
Db 661726 CAAGGCATTATCTCAGCATGCTGTGGCTGGCAATGCTTATTTGTAAGCGGTATAT 661667
Qy 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrp 100
Db 661666 TATAGATTGGAGCGTCTATCGCTTCGGAGCTGTTCTTCTGAATCTTTTGCTTATCTGG 661607
Qy 101 AlaAlaLeuGlnTyrIleuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120
Db 661606 GCAGCTCTACAGTATTGGATGGCCACTCACCTTGTCCAGGACTCGCTGGGATGTTCTT 661547
Qy 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgGluGluPhe 140
Db 661546 GCTATGGGATGGCCGTAGATGCAATGTTCTTGTATTCGAAAGAAATCCGAGAGGAATTT 661487
Qy 141 LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160
Db 661486 TTATTGCTCAAGTCTTAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGAGGCC 661427
Qy 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180
Db 661426 ATTTTGTATCTAACTTGACTACAGTATTGGGCTCAGCACTTCTTTCTTCCTAGATACA 661367
Qy 181 GlyProIleLysGlyPheAlaLeuThrIleLeuGlyIlePheSerSerMetPheThr 200
Db 661366 GGGCCCTATTAAAGGGTTCCTTTGACATTTAGGAATTTTCTCTTCAATGTTTACG 661307
Qy 201 AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisThrGln 220

```

```

Db 661306 GCTCTTTTTCATGACTAAATTTTCTTCATGCTGTGGATGAATAAGACCCCAACATACACAG 661247
Qy 221 LeuHisMetMetAsnLysPheValGlyIleIleYehiAspPheLeuArgGlyCysLysLys 240
Db 661246 TTGCATATGATGAATAAGTTCGTGGGATGAAGCATGATTTCTTGAGAGGATGCAAAAA 661187
Qy 241 LeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla 260
Db 661186 CTTTGGGCTGTTCTGCAAGTGTTCCTTTAGTTGCGTGTCTCTCGGTTGGAGCC 661127
Qy 261 TrpAsnSerValLeuGlyMetAspPheLysGlyTyrAlaPheThrPheAsnProLys 280
Db 661126 TGGAAATCCGTTTTTGGGAATGGATTTTAAAGAGGGGTATGCTTTTACCTTTAAATCCAAA 661067
Qy 281 GluHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGlu 300
Db 661066 GAGCATGGCATCAGCATGTTGCTCAATGCTGGCAAGTTTGTCATTAATCAAGAA 661007
Qy 301 AlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGlyLysIleLys 320
Db 661006 GCTGCTCTTCTCTAGAGACTTCCGTATTCAACAATTGCGATCTTCAGAAAAAGATCAAA 660947
Qy 321 IleTyrPheSerAspLysAlaLeuSerTyrThr-LysGlnIleArgAlaSerLeuLeuLys 340
Db 660946 ATCTATTTTATGATGAAGCTTTAAGCTATATAAGCAGATACAGCAGCTCTCTCTCTAAA 660887
Qy 340 eLeuThrIleMetSerTrpArg-TyrCysGlyIleValValAtqAsnArgProArgPheL 360
Db 660886 ATTAACGATCATGAGCTGGCGTTAGCTGTGGGATGTTGTCAGAAACAGGCTAGATTTTC 660827
Qy 360 euTyrGlyAsnSerLysArgAsnAlaLys-PheTrpSerLysValSerSerLysLeuSer 379
Db 660826 TCTACGGAACCTCTAAACGAAACCGCAAAATTTTGGTCAAGGTAAAGCAGCAACTATCG 660767
Qy 380 LysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeu 399
Db 660766 AAGAAATGCGTTATCAGGCGACCATCGGCTTTTAGGAGCTTTTGGCAATCATCTTGCTC 660707
Qy 400 TyrValSerLeuArgPheGluTyrAlaPheSerAlaValCysAlaLeuIleHis 419
Db 660706 TATGTGAGTTTGGCGTTGAATGGCAATATGCTTTCAGTGCGGTATTCGCTTTTAAATTCAT 660647
Qy 420 AspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPhePheLeuLysLysIleGln 439
Db 660646 GACCTTTTGGCTACTCTGCGAGTCTTCTTTATAGCACATTTCTTTTGAAGAAATTCAA 660587
Qy 440 IleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThr 459
Db 660586 ATAGATTTGCAAGCCATTGGTGCTTTAATGACTGTATTGGGGTATTTCATTAAACAATACT 660527
Qy 460 LeuIleIlePheAspArgIleArgGluAspArgGlnAlaAsnLeuPheThrProMetHis 479
Db 660526 TTGATCATTTTGTGATGTTTGTGTAAGATCGCAAGCAACCTGTTTACCCCTTATGCTAT 660467
Qy 480 ValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThr 499
Db 660466 GTTTTAGTTAATGATGCCCTTCAAAAGACGTTTACGCCGCGATTAACACAGCTACA 660407
Qy 500 ThrLeuSerValLeuLeuMetLeuPheIleGlyGlySerSerValPheAsnPheAla 519
Db 660406 ACTCTATCAGTTTGTGTTAATGCTTTGTTTATAGCGGCTCCTCTGTCTTTTAAATTTGCA 660347
Qy 520 PheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProPro 539
Db 660346 TTTTATATGACCATAGGGATTTCTTAGGAATTTTATGCTCTCTTTATATTATGACCACT 660287
Qy 540 LeuLeuLeuPheMetValArgLysGluAsnArgSerLys 552
Db 660286 CTGTTGTTGTTTATGTCCTGCTTAAAGAAATCGCTCAAAA 660248

```

RESULT 2

US-09-557-884-1/c

; Sequence 1, Application US/09557884

Patent No. 6506581  
 GENERAL INFORMATION:  
 APPLICANT: Fleischmann et al.  
 TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof  
 NUMBER OF SEQUENCES: 1  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Human Genome Sciences, Inc.  
 STREET: 9410 Key West Avenue  
 CITY: Rockville  
 STATE: MD  
 COUNTRY: USA  
 ZIP: 20850  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3 1/2 inch diskette  
 COMPUTER: Dell Pentium  
 OPERATING SYSTEM: MS DOS v6.22  
 SOFTWARE: ASCII Text  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/557,884  
 FILING DATE: 25-Apr-2000  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/476,102  
 FILING DATE: JUN-5-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Michelle S. Marks  
 REGISTRATION NUMBER: 41,971  
 REFERENCE/DOCKET NUMBER: PB186P3  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 301-309-8504  
 TELEFAX: 301-309-8439  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1830121 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 US-09-557-884-1

Alignment Scores:  
 Pred. No.: 1,37e-74 Length: 1830121  
 Score: 719.00 Matches: 186  
 Percent Similarity: 55.69% Conservative: 127  
 Best Local Similarity: 33.10% Mismatches: 174  
 Query Match: 25.88% Indels: 75  
 DB: 4 Gaps: 18

US-09-868-987-14 (1-552) x US-09-557-884-1 (1-1830121)  
 Qy 13 AsnHisAlaSerValSerGlyLysPhe-----Thr 22  
 Db 268552 AATGTTCAACAATTCAAGGACGTTTGGTTCTTAATTTCCAAATTTACTGGTGTGTATAGC 268493  
 Qy 23 HisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheValPro 42  
 Db 268492 ATTGCGGAGCACATAATCTTTTACCTTATTGAAATCTGGTCATTAAATGACCAATT 268433  
 Qy 43 GluValLeuSerGluGluThrIleSerSerAspLeuGlyLysLysGlnCysThrGlnGly 62  
 Db 268432 CAAATTTGTAAGAAGCGCAAAATTTGGCCCATCATTAGTTCGCCAAACGTAAGCAAGG 268373  
 Qy 63 IleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyTyArg 82  
 Db 268372 ATTAATGCGAGTCTTTGGGAGTATTGCTGTTATTGCCCTTTATGTTTACTACAAA 268313  
 Qy 83 PheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrpAlaAla 102  
 Db 268312 ATGTTTGGTGTGATTCGACGTTTGGCTGTTATTAAATCGTATTATCTGTGGGATTA 268253  
 Qy 103 LeuGlnTyLeu---AspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAla 121

Db 268252 ATGCTATTATTTACCGCGGCACACATTCAATGCCGGTATTGCGGGTATCGTTTAACT 268193  
 Qy 122 MetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPheLeu 141  
 Db 268192 TTAGGTATGTCAGTAGATCGAATGATTGATTTTGAACGTATTAAAGAGAAATTCGT 268133  
 Qy 142 LeuSerGlnSerLeuLysSerValGluLysGlyTyThrLysAlaPheGlyAlaIle 161  
 Db 268132 AATGGTCGTTCATTCAGCAAGCCATTAAATGAAGGTATTAAACGGCGCATTTACTTCTATT 268073  
 Qy 162 PheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuSerAlaPheLeuAspThrGly 181  
 Db 268072 TTTGATGCAAACTTAACCAATCTTAACCGCAATTAITCTATACCGGTAGGAACAGCG 268013  
 Qy 182 ProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThrAla 201  
 Db 268012 CCAATTCAAGGTTTTCGATTACGCTTTCACITTTGGTGTTCGATTCTATGTTTACCGCG 267953  
 Qy 202 LeuPheMetThr-----LysPhe 207  
 Db 267952 ATTACAGGAACTCGCGCATTTAGTTAATGCCCTTTACGCTGGTAAACAACTTAAAAAATTA 267893  
 Qy 208 Phe-PheMetLeuTrpMetAsnLys-----ThrGlnHisThrGlnLeuHisMetMetAs 225  
 Db 267892 TTAATTTAGCGCGGAAATGATGAAACTTTTACAAAGACATAAAGACGACATTTTATCCG 267833  
 Qy 225 nLysPheValGlyIleLys-----HisAspPheLeuArgGlyCysLysLys 240  
 Db 267832 TGAATTCATGGGATAAGTAAAGCTCCGTTCCATTGCAATTTATGAAAGTCGTAATTT 267773  
 Qy 240 sleuTrpAlaValSerGlySerValPheLeuLeuGly-----CysValAlaLeuG1 257  
 Db 267772 GGGTTATATATTATCCGCACATTTTTCATGTTAATTTCTCTATTTTATTATACCAAGG 267713  
 Qy 257 yPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGlyGlyTyAlaPheThrPh 277  
 Db 267712 ATTTAAC---TGG-----GGCTTAGATTTTACTGGTGGAGTGGTATTGTATAC 267668  
 Qy 277 eAsnProLysGluHisGlyIleSerAspValAlaGlnMetArgGlyLysValIleHisLys 297  
 Db 267667 TCACTTCTCGCAGTCC-----GCTAACCTTGAACAAATTCGTAGT-----AA 267626  
 Qy 297 sleuGlnGluAlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerG1 317  
 Db 267625 ACTTCACGAAATGGAATTGAAAGCCCA-----ATTGTACAAACCCACAGGATCGGTTCA 267572  
 Qy 317 uLysIleLysIleTyTyPheSerAspLysAlaLeuSerTyThrLysGlnIleArgAlaSe 337  
 Db 267571 GGATGTGATGATT-----CGTTTACCTGCAAG 267545  
 Qy 337 rleuLeuLysLeuThrIleMetSerTrpArgTyCysGlyIleValValArgAsnArgPr 357  
 Db 267544 TAATAATGATTTCTACCAATT-----GGTGAACACGCTCAAAAGTATGTACAGAATGTAGA 267491  
 Qy 357 oArgPheLeuTyTyGlyAsnSerLysArgAsnAlaLysPheTrpSerLysValSerLys 377  
 Db 267490 TAAAGACATT-----CAAAATTCGACGATTTCAGTTC-----GTTGCCCAAAA 267449  
 Qy 377 sLeuSerLysLysMetArgTyTyGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIle1 397  
 Db 267448 TCTTGGTCAAGAAATTAGCACAAAGTGGGTATATCGGACTTACGACATTTAGCAATGGT 267389  
 Qy 397 eLeuLeuTyTyValSerLeuArgPheGluTrpGlnTyAlaPheSerAlaValCysAlaLe 417  
 Db 267388 GCTTATTTATGGGGTCACGTTTGAATGGCGTTTAGGCTTTGGCAGTATCGCTTCTCT 267329  
 Qy 417 uIleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPheLeuLysLys 437  
 Db 267328 TCGGCACACGCTCAATTATTACGTAGGGGTA-----TTCTCTGCATTACA 267284  
 Qy 437 sileGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyTySerLeuAs 457

Db 267283 AATTGAAATGATCTTACTTCTTTCGCGAGCGATTTTATCTCGTGGTGTACTCCATCAA 267224  
Qy 457 nasnThrLeuIlellePheAspAtqIleArgGluAsp---ArgGlnAlaAsnLeuPheTh 476  
Db 267223 CGATAGTATTGGTATTGACCGGGTTCGTGAAAAATTTCCGAAAAATTTAGACGATTGGA 267164  
Qy 476 rPromethisValLeuAlaAsnAspAlaLeuGlnLysThrPheSerArgThrValMetTh 496  
Db 267163 TAGCATTCATATT---ATTGATATTCTTAAACGCAACTTTATCAAGCACTATCATTTAC 267107  
Qy 496 rThraLThrLeuSerValLeuLeuMetLeuLeuPheleGlyGlySerSerValPh 516  
Db 267106 TTCGGTTACTACATAGTTGTCGTGATGCGCATTTCTTCTTGGTGGTCTCTCCATTC 267047  
Qy 516 eAsnPheAlaPheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrl 536  
Db 267046 TAACTTTTCACTTGCTTACTCGTAGGTATTGGATTGGTACTTATTCCTCGATTITGT 266987  
Qy 536 eAla 537  
Db 266986 TGCC 266983

## RESULT 3

US-09-643-990A-1/c

; Sequence 1, Application US/09643990A

; Patent No. 6528289

## GENERAL INFORMATION:

; APPLICANT: Robert D. Fleischmann

; Mark D. Adams

; Owen White

; Hamilton O. Smith

; J. Craig Venter

; TITLE OF INVENTION: The Haemophilus influenzae Rd Genome, Fragments

; Thereof, and Uses Thereof

; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:

; ADDRESS: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville,

; STATE: MD

; COUNTRY: USA

; ZIP: 20850

## COMPUTER READABLE FORM:

; MEDIUM TYPE: 3 1/2 inch diskette

; COMPUTER: Dell Pentium

; OPERATING SYSTEM: MS DOS v6.22

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/643,990A

; FILING DATE: 23-AUG-2000

; CLASSIFICATION: &lt;Unknown&gt;

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/487,429

; FILING DATE: 1995-06-07

; APPLICATION NUMBER: 08/426,787

; FILING DATE: 1995-04-21

; ATTORNEY/AGENT INFORMATION:

; NAME: Kenley K. Hoover

; REGISTRATION NUMBER: 40,302

; REFERENCE/DOCKET NUMBER: PB186P1C1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 301-610-5790

; TELEFAX: 310-309-8439

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1830121 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-643-990A-1

Alignment Scores: 1.37e-74 Length: 1830121  
Pred. No.: 719,00 Matches: 186  
Score: 55.69% Conservative: 127  
Percent Similarity: 33.10% Mismatches: 174  
Best Local Similarity: 25.88% Indels: 75  
Query Match: 4 Gaps: 18  
DB: 18  
US-09-868-987-14 (1-552) x US-09-643-990A-1 (1-1830121)  
Qy 13 AsnHisAlaSerValSerGlyLysPhe-----Thr 22  
Db 268552 AATGTTCAACAATTAAGGACGTTTGGTTCTTAATTTCCAAATTTACTGGTGTGATAGC 268493  
Qy 23 HisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheValPro 42  
Db 268492 ATTGCGGAAGCACATAATCTTTCTACCTTAATGAAATCTGGTGCATTAATTTGCACCAATT 268433  
Qy 43 GluValLeuSerGluThrIleSerSerAspLeuGlyLysLysGlnCysThrGlnGly 62  
Db 268432 CAAATTCGTGAAGAACGACCAATTTGGCCCATTCATTAGTGGCAAAACGTAGACGAAGG 268373  
Qy 63 IleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyArg 82  
Db 268372 ATTAATCGGAGTCTTTGGGAGTAGTTGCTGTATTATGCCCTTTATGTTTGTACTACAA 268313  
Qy 83 PheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuLeuIleTrpAlaAla 102  
Db 268312 ATGTTTGGTGTGATTGCAAGTTTGGCACTTTGTTGTTTATTAATATCTGTTACTTGTGGGATTA 268253  
Qy 103 LeuGlnTyLeu---AspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAla 121  
Db 268252 ATGTCCTATTTTACCCGGCGGACACTTTCAATGCGGGTATTGGCGGTATCGTTTAACT 268193  
Qy 122 MetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPheLeu 141  
Db 268192 TTAGGTATGTCAGTAGATGCGAATGTATTGATTTTGAACGTATTAAAGAAGAAATTCGT 268133  
Qy 142 LeuSerGlnSerLeuLysLysSerValGluLysGlyTyThrLysAlaPheGlyAlaIle 161  
Db 268132 AATGTCGTTCAATTCAGCAAGCCATTAAAGAGTTATAACGGCGGATTTACTTCTATT 268073  
Qy 162 PheAspSerAsnLeuThrValLeuAlaSerAlaLeuLeuPheLeuAspThrGly 181  
Db 268072 TTTGATGCAAACTTAACCAATCTTAACCGCAATTTATCTATACGGGTAGGAACAGGC 268013  
Qy 182 ProIleLysGlyPheAlaLeuThrLeuLeuGlyIlePheSerSerMetPheThrAla 201  
Db 268012 CCAATTCGAAGTTTGGGATTACGCTTTCACTTGTGTGCGATTCTATGTTTACCCTG 267953  
Qy 202 LeuPheMetThr-----LysPhe 207  
Db 267952 ATTACAGGAATCGCGCATTTAGTTAATGCCCTTTACGGTGTAAACAACTTAAATAAATTA 267893  
Qy 208 Phe-PheMetLeuTrpMetAsnLys-----ThrGlnHisThrGlnLeuHisMetMetAs 225  
Db 267892 TTAATTTAGCGGGAATGATGAAACTTTTTACAAAAGATAAAGACGACATTTTATCCG 267833  
Qy 225 nLysPheValGlyIleLys-----HisAspPheLeuArgGlyCysLysLys 240  
Db 267832 TGAATCAATGGGATAAAGCTCCCGTTCCCATTCACCTGAATTTATGAAAGTCGCGTAAAT 267773  
Qy 240 sLeuTrpAlaValSerGlySerValPheLeuLeuGly-----CysValAlaLeuGly 257  
Db 267772 GGGTTATATATATCCGCACTTTTGTGATGGTAATTTCTCTATTTTATTATACCAGG 267713  
Qy 257 yPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGlyGlyTyAlaPheThrPh 277  
Db 267712 ATTTAAC---TGG-----GGCTTAGATTTTACTGCGAGTGGTATTGTATAC 267668  
Qy 277 eAsnProLysGluHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLys 297  
Db 267667 TCACTTTCGCGAGTCC-----GCTAACCTTTGAACAAATTCGTAGT-----AA 267626

|    |        |   |        |
|----|--------|---|--------|
| Qy | 297    | sLeuGlnGluAlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerG   | 317    |
| Db | 267625 | ACTTCACGAAATGGAATTGAAAGCCCA-----ATTGTACAACACACAGGATCGGTCA     | 267572 |
| Qy | 317    | uLysIleLysIleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSe  | 337    |
| Db | 267571 | GGATGTGATGATT-----CGTTTACCTGCAAG                              | 267545 |
| Qy | 337    | rLeuLeuLysLeuThrIleMetSerTrpArgTyrCysGlyIleValValArgAsnArgPr  | 357    |
| Db | 267544 | TAATTAATGATCTACCAATT-----GGTGAACACGCTCAAAAGTATGCTACAGAAATGAGA | 267491 |
| Qy | 357    | oArgPheLeuTyrGlyAsnSerLysArgAsnAlaLysPheTrpSerLysValSerSerLys | 377    |
| Db | 267490 | TAAAGACACATT-----CAAAATTCGACGATTGAGTTC-----GTTGGCCCAAA        | 267449 |
| Qy | 377    | sLeuSerLysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleIl  | 397    |
| Db | 267448 | TGTTGGTGAAGAATTAGCACACAGGTGCGGTATATCGCACTTTAGCGACATTAGCAATGGT | 267389 |
| Qy | 397    | eLeuLeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaIle | 417    |
| Db | 267398 | GCTTATTTATGTGGGTCACGTTTTGAATGGCGTTTAGGCTTTGGCAGTATCGCTTCTCT   | 267329 |
| Qy | 417    | uIleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPheLeuLysLys    | 437    |
| Db | 267328 | TGCGCAGCAGCTCATTTATTCGTTAGGGGTA-----TTCTCTGCATTACA            | 267284 |
| Qy | 437    | sIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAs  | 457    |
| Db | 267283 | AATTGAAATTCATCTTACTTTGTGCGACGCAATTTATCTGTGTGGGTACTCCATCAA     | 267224 |
| Qy | 457    | nAsnThrLeuIleIlePheAspArgIleArgGluasp--ArgGlnAlaAsnLeuPheTh   | 476    |
| Db | 267223 | CGATAGTATTGTGGTATTGACCGGTTTCGTAATAATTCGAAATAATAGACGATTGGA     | 267164 |
| Qy | 476    | rProMethHisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetTh | 496    |
| Db | 267163 | TACGATTGATATT--ATTGATATTTCTTTAACGCAAACTTTATCAGAACTATCATTC     | 267107 |
| Qy | 496    | rThrAlaThrThrLeuSerValLeuLeuMetLeuPheIleGlyGlySerSerValPh     | 516    |
| Db | 267106 | TTCCGTTACTACATTAGTTGCGATTGGCAATGCTTCTTCTTGGTGGTCTCTCCATTCA    | 267047 |
| Qy | 516    | eAsnPheAlaPheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIl  | 536    |
| Db | 267046 | TAACCTTTTCACTGCTTTACTCGTAGGATTGGATTGGTACTTATTCTCCGATTTTTGT    | 266987 |
| Qy | 536    | eAla  | 537    |
| Db | 266986 | TGCC  | 266983 |

```

RESULT 4
US-09-462-844-1
; Sequence 1, Application US/09462844
; Patent No. 6258563
; GENERAL INFORMATION:
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: Increasing Production of Proteins in
; TITLE OF INVENTION: Gram-Positive Microorganisms
; FILE REFERENCE: GC385-US
; CURRENT APPLICATION NUMBER: US/09/462,844
; CURRENT FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: PCT/US98/14786
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: EP 97305286.3
; PRIOR FILING DATE: 1997-07-16
; PRIOR APPLICATION NUMBER: EP 97305344.0
; PRIOR FILING DATE: 1997-07-17
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1

```

|      |   |      |
|------|---|------|
| 1399 | GGGATTATTATCTCGTCTGTGTTTCAGCGCTGAAT-----CTTGGCATTGACTTTGCAAGC   | 1455 |
| 272  | Gly-----TyrAlaPheThrPheAsnProLysGluHisGly                       | 283  |
| 1453 | GGTGACGGATTGAAGTGCAAACCGACCATACCTGCAGCAGCAAGTTGAGAAGGAT         | 1512 |
| 284  | IleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGluAlaGlyLeu    | 303  |
| 1513 | TTTGAATCTCTGGGTATGACCCCTGATCTACTGTAGTTCTGTTCAGCGCAAAAGACAATATC  | 1572 |
| 304  | Ser---SerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLysIleTyr    | 322  |
| 1573 | GGGTGGCCCGTTTTGTGGGGTCCGAGATAAAGAAACCAATTCGAAAGATAAAACGATAT     | 1632 |
| 323  | PheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLysLeuThr       | 342  |
| 1633 | TTTAAGACDAA-----  | 1644 |
| 343  | IleMetSerTrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeuTyrGly    | 362  |
| 1645 | -----TACGGA   | 1650 |
| 363  | AsnSerLysArgAsnAlaLysPheTrpSerLysValSerSerLysLysLeuSerLysLysMet | 382  |
| 1651 | TCTGATCCAAATGTCAGCACAGTTCACCAGCAGTCGGTAAGGAGTGGCGGCAAAATGCG     | 1710 |
| 383  | ArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleIleLeuLeuTyrValSer    | 402  |
| 1711 | CTGTAGCAGCTTGCTATAGCTTCTATTGGC-----ATCATTATTTAGTTTCA            | 1758 |
| 403  | LeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeu    | 422  |
| 1759 | ATCCGATTCCGAATACAAAATGGCGATTGTCGCCATCGCCTCAITGCTATATGAC-----    | 1812 |
| 423  | AlaThrCysAlaValLeuPheIleAlaHisPhePheLeuLysLysIleGlnIleAspLeu    | 442  |
| 1813 | -----GCATTCTTTATCGTCACGCTCTCTCAGTATTACAAAGGCTTGAGGTAGTAGTT      | 1863 |
| 443  | GlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIle    | 462  |
| 1864 | ACATTATCGGGCCATCTTGACGATTAATCGGGTATTCATTAACGATACAAATCGTTACA     | 1923 |
| 463  | PheAspArgIleArgGlu-----AspArgGlnAlaAsnLeuPheThrProMetHis        | 479  |
| 1924 | TTTGACAGGCTCCGCGAGCATATGAAAAGCGTAAGCCGAAACCTTTGCCGATCTCAAC      | 1983 |
| 480  | ValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThr    | 499  |
| 1984 | CATATTGAACCTTGACCCCTGCAGCAACCTTTACAGCTTCAATTAACATGTATTAAAC      | 2043 |
| 500  | ThrLeuSerValLeuLeuMetLeuLeuPheIleGlyGlySerSerValPheAsnPheAla    | 519  |
| 2044 | GTGTGTGATTGTGTGTGACATTCCTGATCTTTGGAGCATCTTCTATTACTAACTCTCA      | 2103 |
| 520  | PheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProPro    | 539  |
| 2104 | ATTGCTTTATGGTCGGCTGTTAAACAGCGCTTTATTCTTCTCTATACATTGCCGCAA       | 2163 |
| 540  | LeuLeuLeuPheMetValArgLysGluAsnArgSerLys                         | 552  |
| 2164 | ATTTGGCTTGC-ATGCGAAAGCAAGACATCGAAAGAA                           | 2201 |

## RESULT 5

US-09-221-017B-311/c  
; Sequence 311, Application US/09221017B  
; Patent No. 6444799  
; GENERAL INFORMATION:  
; APPLICANT: ROSS, Bruce C.  
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
; NUMBER OF SEQUENCES: 1120  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: MORRISON & FOERSTER  
; STREET: 755 PAGE MILL ROAD

|    |      |   |      |
|----|------|---|------|
| Db | 1875 | GGATTCTTCTGCTCTCTGCTCGCTTGGTTATCTCTGATGTGTTACATGTGTCTGCCTTAC  | 1816 |
| QY | 82   | ArgPhe---GlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuLeuTrp     | 100  |
| Db | 1815 | GGTTTCTTCCGGGTCTTATCGAAACGGCGCATGATTGTTAAACAGCTTCTTCACATG     | 1756 |
| QY | 101  | AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu  | 120  |
| Db | 1755 | GGCGTATTGGCTTCTTCCATGCGGTCTGTACCCCTCTCGGGTATCGCAGGTGTGGTGTG   | 1696 |
| QY | 121  | AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArGluGluPhe   | 140  |
| Db | 1695 | ACGTGGGTATGGCTGTGGATGCCAACGTACTTATCTTCGAGCGTATCAAGAAGAGCTT    | 1636 |
| QY | 141  | LeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGlyAla  | 160  |
| Db | 1635 | CGTGGGTAGACTCCGATTCGTGGCGGTACGGATGTTATGCCAACGCTTCTCTGCC       | 1576 |
| QY | 161  | IlePheAspSerAsnLeuThrValLeuAlaSerAlaLeuPhePheLeuAspThr        | 180  |
| Db | 1575 | ATCTTCGACTCGAAGTTACGACTATTATTACCGGTATCATCTTCTTCACGGGACG       | 1516 |
| QY | 181  | GlyProIleLysGlyPheAlaLeuThrIleuLeuGlyIlePheSerSerMetPheThr    | 200  |
| Db | 1515 | GGGCGGATTCGCGGTTTGGCACATCGTGTGATTATCGGTCTTATCGCTTCTTTCATTACG  | 1456 |
| QY | 201  | AlaLeuPheMetThrLysPhePhePhe-MetLeuTrpMetAsnLysThrGlnHisThrG   | 220  |
| Db | 1455 | GCTGTCTTCTTGACTCGTATCGTCTTCCAGAACTGGCGAAAGAGTCTGTTGGATAAG     | 1396 |
| QY | 220  | nLeuHisMetMetAsnLysPheValGlyIleLys-----HisAspPheLeuAr         | 236  |
| Db | 1395 | ATTACATTCACTACGACGATTACTCGCAATCTCTTGTCAATCCCTCATACAAC-ATCTT   | 1337 |
| QY | 236  | gGlyCysLysLysLeu---TrpAlaValSerGlySerValPheLeuLeuGlyCysValAl  | 255  |
| Db | 1336 | GGGTAAGCGCAAGACCGGCTTTATCATCTCGGTGATTATCATCGTTTGGGACTTATAGC   | 1277 |
| QY | 255  | a-----LeuGlyPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGlyG       | 272  |
| Db | 1276 | TTCATTACAACTCGTCTCAATAGG-----GGTATTGAATCTTCGGAGG              | 1232 |
| QY | 272  | Y-----TyrAlaPheThrPheAsnProLysGluHisGlyIleSerAspValAlaGlnMe   | 290  |
| Db | 1231 | ACGTAACACTACGTAGTTAAATTC-----GACCAGCGGTATCTCCGAAGCC---GT      | 1184 |
| QY | 290  | tArgGlyLysValValHisLysLeuGlnGlu-----AlaGlyLeuSe               | 304  |
| Db | 1183 | TCGTTCGGCTTGTCTTCTCCCTCGCAGGAAAGGATTGGTTACCTCCATCGGTACTGA     | 1124 |
| QY | 304  | rSerArgAspPheArgIleGlnThr-----PheGlySerSerGluLysIleLys        | 320  |
| Db | 1123 | AGGCAGAGGTGGTATATCTACGAACCTATAGATCCAGGAGGAAAGCAAGAACTGA       | 1064 |
| QY | 320  | sIleTyrPheSerAspLysAlaLeuSer-----TyrThrLysGlnIleAr            | 335  |
| Db | 1063 | AGCAGAGATTACTGACAAATTTGATCAGACCCCTGAAAGGTTTCTACACCCAGCAGCTTAC | 1004 |
| QY | 335  | gAlaSerLeuLeuLysLeuThrIleMetSerTrpArgTyrCysGlyIleValValArgAs  | 355  |
| Db | 1003 | TGCTGATCAGTCTCTTGACAATATCATTTAGCTCTCAG-----                   | 967  |
| QY | 355  | nArgProArgPheLeuTyrGlyAsnSerLysArgAsnAlaLysPheTrpSerLysValSe  | 375  |
| Db | 966  | -----AAAGTAAG   | 959  |
| QY | 375  | rSerLysLeuSerLysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAl  | 395  |
| Db | 958  | TCCAGTATGTCAGTGACATCAGAGGTGCTATTTGGGCTGTGCTGTATCATCAT         | 899  |
| QY | 395  | alleIleLeuLeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPheSerAla----   | 413  |
| Db | 898  | CTTCATGGCCATTATATTCTGATTCGCTTCGT---GACATTCTTCTCTCGCGGGT       | 842  |

QY 414 -----ValCysAlaLeuHisAspLeuLeuLeuAlaThrCysAlaValLeuPhe-- 429  
Db 841 ATTCTATCTGTGGCGCTACTACATTCTGCATTATTCTCTGTATCGTTCGTGGAA 782  
QY 430 -IleAlaHisPheLeuLysLysIleGlnIleAspIleuGlnAlaIleGlyAlaLeuMe 449  
Db 781 GATTCTGCCCTTC-----ACCATGGAGATCGATCAGAACTTCATCGCTGCTATTCT 731  
QY 449 tThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIlePheAspArgIleArgGluAs 469  
Db 730 GGCTATCATCGTTACTCGCTCAATGACACCGTGGTGTATTGTGACCGTATCCGAGAGAC 671  
QY 469 pArgGlnAlaAsnLeuPheThrProMetHisValLeuValAsnAspAlaLeuGlnLysTh 489  
Db 670 GATGAAATGTACCCCAACAGAGATCGCTATCAGGTGATCAACGATGCCCTTAATTCAAC 611  
QY 489 rPheSerArgThrValMetThrAlaThrAlaThrLeuSerValLeuLeuMetLeuLeuPh 509  
Db 610 ATTGGGTGCAACATTAAATACGTCCTTGACTACGTTTATCGTTATGTTGGTAATCTTCAT 551  
QY 509 eileGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeuGl 529  
Db 550 CTTTGGAGGTCTACGATGCGTAGTTTCACGTTCTCGATCCTGCTCGGTATCGTTATCGG 491  
QY 529 yThrLeuSerSerLeuTyrIleAlaProProLeuLeuLeuPheMetValArgLysGluAs 549  
Db 490 TACATCTCTACGCTCTTTGTTGCTACACCC---CTTGSCCTACGAGATCCAAAGCGCAA 434  
QY 549 nArgSerLys 552  
Db 433 GCTCAACAAA 424  
RESULT 6  
US-09-134-001C-852  
; Sequence 852, Application US/09134001C  
; Patent NO. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHY  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GPC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 852  
; LENGTH: 2301  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-852

|                        |          |
|------------------------|----------|
| Alignment Scores:      |          |
| Pred. No.:             | 1.62e-70 |
| Score:                 | 646.50   |
| Length:                | 2301     |
| Matches:               | 164      |
| Conservative:          | 129      |
| Percent Similarity:    | 51.22%   |
| Best Local Similarity: | 28.67%   |
| Mismatches:            | 200      |
| Query Match:           | 23.27%   |
| Indels:                | 79       |
| DB:                    | 4        |
| Gaps:                  | 12       |

US-09-868-987-14 (1-552) x US-09-134-001C-852 (1-2301)

|    |     |  |     |
|----|-----|--|-----|
| Qy | 2   | ValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLysPhe | 21  |
|    |     |  |     |
| Db | 625 | GTAGACCAACCTATTAATTCCTAGTGTGAAATTTCAGTGCCTTCATGGCAAAAAA      | 684 |
|    |     |  |     |
| Qy | 22  | ThrHisArgLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheVal          | 41  |
|    |     |  |     |
| Db | 685 | GGTGTTGAAGAAGCGCAACAAATAGCTGAGTTATTAATCCCGGCTCAATACCACTTGAT  | 744 |
|    |     |  |     |





Db 2910893 CTGGACTCGCAGGTTGTCAAGTCCAGGATCCAGGAGCGATCCCGCGCGCGAACC 2910834  
QY 21 -----PheThrHisArgGluValSerLysLeuAlaSerAsp 32  
Db 2910833 CAGATCAGCGGTGGGATCCCGCAATTCACCGCGGACCGACCGAGCTCGCAACGTC 2910774  
QY 33 LeuLysSerGlyAlaMetSerPheValProGluValLeuSerGluGluThrIleSerSer 52  
Db 2910773 CTCAAGTACGGGTGCTCGCTCCCTGCTCCGAAACGTCGGAAGCCAAACGGTTTCGGCG 2910714  
QY 53 AspleuGlyLysLysGlnCysThrGlnGlyIleIleSerAlaCysCysGlyLeuAlaMet 72  
Db 2910713 ACCTTGGGTTGTATCGCTCGCGCGGCGATGATCGGGTGCATCGCGCTCTGTTG 2910654  
QY 73 LeuIleValLeuMetSerValTyrThrArgPheGlyGly-----Val 86  
Db 2910653 GTGTGGTGTATTCGCTCTACTACCGAGTGTGGGACTGCTCAGCGCTTTCGCTA 2910594  
QY 87 IleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrpAlaAlaLeuGlnTyrLeu 106  
Db 2910593 GTGGCTTCGCTCAATGGTGTTCGATCTGCTGGGCTCTTGGGC-----CGATATATC 2910540  
QY 107 AspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAlaMetGlyMetAlaVal 126  
Db 2910539 AACTACACCTTAGATCTGGCTGGTATTGCGGCTGATCATCGGTATCGGACCAACCGCC 2910480  
QY 127 AspAlaAsnValLeuValPheGluArgIleArgGluGluPheLeuLeuSerGlnSerLeu 146  
Db 2910479 GACTCGTTCGGTGGTTCGAAACGATCAACAGCAGATCCCGAAGCGGTTCGTTTC 2910420  
QY 147 LysLysSerValGluLysGlyTyrThrLysAlaPheGlyAlaIlePheAspSerAsnLeu 166  
Db 2910419 CGGTCCGAGTCCCGCGGTTGGCGGTGCCCGAGAGATCGTGTGGGCAACGCC 2910360  
QY 167 ThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGlyProIleLysGlyPhe 186  
Db 2910359 GTCACTTCTCGCGCGTGGCTGCTGTACTTTCTGGCGATCGGTCAAGTGAAGGGTTC 2910300  
QY 187 AlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThrAlaLeuPheMetThrLys 206  
Db 2910299 GCGTTCACCTCG-----GGCTCACACGATCCTCGGACCTCGTCGTG 2910258  
QY 207 PhePhePheMetLeuTrp-----MetAsnLys 215  
Db 2910257 GTGTCTTGTGAGTGGCGGTGGTGTATCTGGCTGCCAAATCGTCTGCTGGCCAAAG 2910198  
QY 216 ThrGlnHisThrGlnLeuHisMetMetAsnLysPheValGlyIleLys-----231  
Db 2910197 CCGGCATACAACGGCTGGGAGCGGTGCAGCAGGTGCACGCGAAACCGCGGCATGGCT 2910138  
QY 231 -----231  
Db 2910137 CGGACGGGACGGGATAGCCGATGGCGTCAAAAGCAAGACCGCGCGACGACGAAGCC 2910078  
QY 231 -----231  
Db 2910077 ACGTGCGGGTGCNACTCACCGAGGCCACCGAGAGTCTGTGGCGGACCGACGGCGAC 2910018  
QY 232 -----HisAspPheLeuArg-----236  
Db 2910017 AGCACACAGACACCGCTCGAAGCTGGGGCACCATTAGCTTCCTTTCGCGCTGTACACC 2909958  
QY 237 -----GlyCysLysLysLeuTrp---AlaValSerGlySer 247  
Db 2909957 GGCACCGGGCGTTTGTAGGTAGTGGGACCGCGGTTGTGGTGTGGCGTGGCGGTGGC 2909898  
QY 248 ValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerValLeuGlyMet 267  
Db 2909897 ATCGTCGGGTGGCCATCGCCAGCATGCTTTCTGGGGTTC-----ACCTTCGGCATC 2909844  
QY 268 AspPheLysGlyGlyTyrAlaPheThrPheAsnProLysGluHisGlyIleSerAspVal 287  
Db 2909843 GATTTCAAAGCGGTACACCGGTGTCTGTTTC-----CCGCGTGGCAGACCCCAAGTC 2909793

QY 288 AlaGlnMetArgGlyLysValValHisLysLeuGlnGluAlaGlyLeuSerSerArgAsp 307  
Db 2909792 GCGCAGGTCAAGACGCTCTACTACCGGGCGCTC-----GGCAGTGAAGCCCGCATCG 2909742  
QY 308 PheArgIleGlnThrPheGlySerSerGluLysIleLysIleTyrPheSerAspLysAla 327  
Db 2909741 GTGGTATCTCGGGCGGTGCTTCGGGACCGTGCAGATTCTGTCGGAGACGCTGACC 2909682  
QY 328 LeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeuThrIleMetSerTrpArg 347  
Db 2909681 AGCGACAGACGCGCAAGCTGGGGAGCGCTTATTCTGAGGCCTTC-----2909637  
QY 348 TyrCysGlyIleValValArgAsnArgProArgPheLeuTyrGlyAsnSerLysArgAsn 367  
Db 2909636 -----GGGCCCCAAGGAACGAGCGTCAGCCCGCAGCAAGCAG 2909601  
QY 368 AlaLysPheTrpSerLysValSerSerLysLysMetArgTyrGlnAlaThr 387  
Db 2909600 GCATCAGGACTCGCGGTGTTCGGAGACCTGGGGCGGTTCAGATCACCAAGAAAGCGGTG 2909541  
QY 388 IleGlyLeuLeuGlyAlaLeuAlaIleIleLeuLeuTyrValSerLeuArgPheGluTrp 407  
Db 2909540 ATCGCGCTGGTGTGTTCTGTGTGTGGTGGCTCTCTATACCTGCGCTACGAGCGC 2909481  
QY 408 GlnTyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaVal 427  
Db 2909480 TACATGACCATCTCGCGGATAACGGCATGCTCTTCGACCTGACCGTCACCGCGCGTG 2909421  
QY 428 LeuPheIleAlaHisPheLeuLysIleGlnIleAspLeuGlnAlaIleGlyAla 447  
Db 2909420 TATTCCTCGTGGTTC-----GAAGTCACCCCGCGCCACGGTCATCGGC 2909376  
QY 448 LeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIlePheAspArgIleArg 467  
Db 2909375 CTGCTGACCATCTCGGGTTCCTCGCTCTATGACACCGTCTATGTTGCACAAGGTTCGAG 2909316  
QY 468 GluAsp-----ArgGlnAlaAsnLeu 474  
Db 2909315 GAGAACACCATCTGGCTTCAGCACACACCCCGCGCACCTTCGCCGAGCAGGCCAACCTG 2909256  
QY 475 PheThrProMetHisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrVal 494  
Db 2909255 -----GGCATCAACCAACGCTTCATGCGTGCATC 2909226  
QY 495 MetThrThrAlaThr---ThrLeuSerValLeuLeuMetLeuLeu-----Phe 509  
Db 2909225 AACACACCGCTCATCGCGGTGTTCGGGTGTTCGGCGTGTATGCTGTGGCGGTCTGGCTG 2909166  
QY 510 IleGlyGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeuGly 529  
Db 2909165 CTGGGCGTCCGACCTCTCAAAGACCTGGCGCTGGTGGAGCTGATCGGCATCATCGCC 2909106  
QY 530 ThrLeuSerSerLeuTyrIleAlaProProLeuLeuLeuPhe-----543  
Db 2909105 ACCTACTGCTCAATCTTCTTCGCCACCGCGTGTGGT-TACCTCGTGAGCGCAGGAG 2909046  
QY 544 MetValArgLysGluAsnArg 550  
Db 2909045 TTGGTGGCGCAACCAACCGCT 2909025

RESULT 8  
US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00





Db 12513 AGGCGGCACACCGGTGTTCATG---CCTGTCTCGCCCGAGTGGGGGGACAGCGCAA 12457  
Qy 290 etArgGlyValValHisLysLeuGlnGluAlaGlyLeuSerSerArgAspPheArgI 310  
Db 12456 TTGAGTTGGCAGGTTCGCCATGCTTCAAGAGACCCCTTGGCAGTATCCGGAATCGG 12397  
Qy 310 leGlnThrPheGlySerGlyLysIleLeuLysIleTyPheSerAspLysAlaLeuSer- 329  
Db 12396 TGGTCGTTGGTAACCGCGGTGACCAACCGTGGCATCAGTTCGAAACCGTGTCCA 12337  
Qy 330 -----TyThrLysGlnIleArgAlaSerLeuLeuLysLeuThrIleMetSerTrpArgT 348  
Db 12336 ATGACACAGCTGCAAGCTGCGTACGCTTGTGACGCATTC----- 12293  
Qy 348 yrCysGlyIleValValArgAsnArgProArgPheLeuTyGlyAsnSerLysArgAsnA 368  
Db 12292 -----GGACCCAAAGGCGCGAGCTAAGCTTAGCAAGCAGG 12256  
Qy 368 laLysPheTrpSerLysValSerLysLeuSerLysLysMetArgTyGlnAlaThrI 388  
Db 12255 CAATCAGCAGCGCGGTATCGGAGACCTGGGGCGGTGAGTACCAAGAGGTGTGA 12196  
Qy 388 leGlyLeuLeuGlyAlaLeuAlaIleLeuLeuTyValSerLeuArgPheGluTrpG 408  
Db 12195 TCGCTCTTGTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 12136  
Qy 408 InTyAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValL 428  
Db 12135 ATATGGCTATCTCCGACGTACTACATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 12076  
Qy 428 euPheIleAlaHisPheLeuLysIleGlnIleAspLeuGlnAlaIleGlyAlaL 448  
Db 12075 ACTCGCTGTGGCTTC-----GAAGTCACTCCGCGCACTGTCTCGCT 12031  
Qy 448 euMetThrValLeuGlyTySerLeuAsnAsnThrLeuIlePheAspArgIleArgG 468  
Db 12030 TGTTCACGATCTCGGTTCTCTCTCTACGACACCGTTATCGTGTGTGTGTGTGTGT 11971  
Qy 468 luAsp-----ArgGlnAlaAsnLeuP 475  
Db 11970 AGAATACCCACGGTTTCCAGCACACCACTCGCGGCACCTTCGCTGAGCAGGCTAATTTG- 11912  
Qy 475 heThrProMetHisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValM 495  
Db 11911 -----GCGATCAATCAGACCTTCATGCGTTCGATCA 11881  
Qy 495 etThrAlaThrThrLeuSerValLeuLeuMetLeuLeu-----PheI 510  
Db 11880 ACACCGCCTGATCTCGGTACTACCGGTGTGTGGCGCTGATGTGTGTGTGTGTGTGT 11821  
Qy 510 leGlyGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeuGlyT 530  
Db 11820 TGGGCGTGGTACGCTCAAGGACCTGGCATTAAGTGCAGCTGTGTGGCATATTGTGCGCA 11761  
Qy 530 hrLeuSerSerLeuTyIleAlaProProLeuLeuLeu 542  
Db 11760 CTAATCGTGTGATCTCTTTGCAACTCCACTGTGTGTGTC 11723

## RESULT 10

US-09-252-991A-6413  
; Sequence 6413, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,789  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6413  
; LENGTH: 1398  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6413  
Alignment Scores:  
Pred. No.: 4,46e-43 Length: 1398  
Score: 425.00 Matches: 96  
Percent Similarity: 59.09% Conservative: 47  
Best Local Similarity: 39.67% Mismatches: 79  
Query Match: 15.30% Indels: 20  
DB: 4 Gaps: 3  
US-09-868-987-14 (1-552) x US-09-252-991A-6413 (1-1398)  
Qy 1 MetValSerSerProLeuLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20  
Db 727 ATCATCAGCTGGCGACCATCCAGTCGCCCTGGCAACACAGTTCCGTATATCACCGGCTC 786  
Qy 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40  
Db 787 GATGGCCCCGGCGAGTCTCCGAGCTGGCCCTCTGCTGCTGCGCGCGCTGCTGCGG 846  
Qy 41 ValProGluValLeuSerGluThrIleSerSerAspLeuGlyLysGlnCysThr 60  
Db 847 CCATGTACTTCCCGGAGAACCGACCATCGGCCGAGCTGGGGCGGCAACATCGCC 906  
Qy 61 GlnGlyIleLeuSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTy 80  
Db 907 AAGGTATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 966  
Qy 81 TyArgPheGlyGlyValIleAlaSerGlyValAlaValLeuLeuLeuLeuLeuLeuLeu 100  
Db 967 TACCGTTTCTTCGCGGTGATCGGCACCGTGGCTGGCTTCAACATGGTTCATGCTGCTG 1026  
Qy 101 AlaAlaLeuGlnTyLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120  
Db 1027 GCGCTGATGTGATCTCTCGCGGCGACCTGACCTCGCGGCGATCGCGGTATCGTCTG 1086  
Qy 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluPhe 140  
Db 1087 ACCATGGGCATGGCGGTGGAGCAACGTCGTGATCTTCTCGCGGATACGCGGAACTG 1146  
Qy 141 LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyThrLysAlaPheGlyAla 160  
Db 1147 GCCAATGGCATGTCCGTCCTCAAGAGCGATCCAGAGGCTTCAACCGTGTCTTCCCGCG 1206  
Qy 161 IlePheAspSerAsnLeuThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180  
Db 1207 ATCTCGACGCCAACCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1266  
Qy 181 GlyProIleLysGlyPheAlaLeuThrLeuLeuLeuGlyIlePheSerSerMetPheThr 200  
Db 1267 GGGCGGTGAAGATTGGCGTGACCATGTCGTGCGCATCATCATCATCATCATCATCATCAT 1326  
Qy 201 AlaLeuPheMetThrLysPhePhePheMetLeuTrpMetAsnLysThrGlnHisThrGln 220  
Db 1327 GCCATGTGTCAACCGC----- 1344  
Qy 221 LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysValLys 240  
Db 1345 ---GCAATGTGTCAACCTGATCTTCGGC---GGGCGTGACTTC-----AAGAAG 1386  
Qy 241 LeuTrp 242  
Db 1387 CTGTGG 1392  
RESULT 11  
US-09-328-352-1027  
; Sequence 1027, Application US/09328352  
; Patent No. 6562958

## ; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GT99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 1027  
; LENGTH: 1947  
; TYPE: DNA  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-1027

## Alignment Scores:

Pred. No.: 6e-42 Length: 1947  
Score: 418.00 Matches: 85  
Percent Similarity: 64.68% Conservative: 56  
Best Local Similarity: 38.99% Mismatches: 77  
Query Match: 15.05% Indels: 0  
DB: 4 Gaps: 0

US-09-868-987-14 (1-552) x US-09-328-352-1027 (1-1947)

QY 1 MetValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20  
Db 1273 GTGATTAATGTCGCAACTGTTCAAGCCGTTTAGGTTCAACTTTCCTGATATCTGTTTA 1332  
QY 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40  
Db 1333 AGTTCACTCAAGAAGCATCTGAACCTTCGCTTAATGCTGCTGCTGCTGCTGCTGCTG 1392  
QY 41 ValProGluValLeuSerGluThrIleSerSerAspLeuGlyLysGlnCysThr 60  
Db 1393 CCAATGTACTTCGTGAAGACGCTGTTGGTCCAAAGCCTTGGTCAAGAAACATGTAT 1452  
QY 61 GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80  
Db 1453 AAAGGTGTACTATCAACGCAAAATCGGTTCTTCTGTTGTCATCTGGATGCTGATTC 1512  
QY 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuLeuLeuLeuLeuLeuTrp 100  
Db 1513 TTCCGTCATTGCTTTAAATCGCAACTTTGCACCTTGTATTTAACTTGGCAATGATTTTA 1572  
QY 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120  
Db 1573 ACCATATGTCTGGATAGTGCTTCCTCACCTTACCGGGTATTCGGGTATCGTCATT 1632  
QY 121 AlaMetGlyMetAlaValAspAlaAsnValPheGluArgIleArgGluGluPhe 140  
Db 1633 ACCATTGGTATGGCGTCGACGCAACGTACTGATGTGAACGTATTCGAGAAGAAATG 1692  
QY 141 LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160  
Db 1693 CTCTGGGGGCGCTCACCTAAACAGGCCATTTGTGGCGGGTTATGATCGAGCTATAACACC 1752  
QY 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPheLeuAspThr 180  
Db 1753 ATTTTCGACTCGAACTTAAACAGCTTCCTTGTGGTCATTCAATCTGTTTGAATCGGTACA 1812  
QY 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200  
Db 1813 GGCCCGATTAAAGGTTTCGCGGACATTAATGATCGGTATTTATTTGCTCAATGTTTACT 1872  
QY 201 AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHis 218  
Db 1873 GCCATTACAGTAACGCGTGGATTGTACAAATCATTTATGGTAACGCGGTAAC 1926

## RESULT 12

US-09-252-991A-6330  
; Sequence 6330, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6330  
; LENGTH: 1890  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6330

## Alignment Scores:

Pred. No.: 7.61e-42 Length: 1890  
Score: 417.00 Matches: 87  
Percent Similarity: 64.56% Conservative: 46  
Best Local Similarity: 42.23% Mismatches: 73  
Query Match: 15.01% Indels: 0  
DB: 4 Gaps: 0

US-09-868-987-14 (1-552) x US-09-252-991A-6330 (1-1890)

QY 1 MetValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20  
Db 1241 ATCATCAGCTGGCGACCATTCAGTCGCGCTGGCGAACAGTTCGTTATCATCCGCGCTC 1300  
QY 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40  
Db 1301 GATGCGCGCGGAGTCTCCGAGTGGCCCTGCTGCTGCGCGCGCGCTGCTGCG 1360  
QY 41 ValProGluValLeuSerGluThrIleSerSerAspLeuGlyLysGlnCysThr 60  
Db 1361 CCGATGTACTTCGCGAAGAACGACCATTCGCGCGAGCTGGCGCGCGGCGGACATCGCC 1420  
QY 61 GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80  
Db 1421 AAGGCTATCGATCGCTCTGGGCGATGCTCTTCGTATCGCTTTCATCATCATCATC 1480  
QY 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuLeuLeuLeuLeuLeu 100  
Db 1481 TACCGTTTCTCGCGGTGATCGCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1540  
QY 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120  
Db 1541 GCGCTGATGTCGATCTTCGCGCGACCTTGACCTTCGCGCGGATCGCGGTATCGTGTG 1600  
QY 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPhe 140  
Db 1601 ACCATGGCGATGGCGGTGGAGCGCAACGCTGCTGATCTTCTCGCGGATACGCGGAACTG 1660  
QY 141 LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160  
Db 1661 GCCAATGGCATGTCGCTCCAAAGCGCATCCACGAAGGCTTCAACCGGCTTCAACCGG 1720  
QY 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPheLeuAspThr 180  
Db 1721 ATCTCTGACGCGCAACCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1780  
QY 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200  
Db 1781 GGCCCGGTGAAGGATTTGCGGTGACCATGCTGCTGCGCATCATCACCTCGATGTTTACG 1840  
QY 201 AlaLeuPheMetThrLys 206  
Db 1841 GCCATCATGGTACCCCGC 1858

## RESULT 13

US-09-252-991A-6245

```

; Sequence 6245, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6245
; LENGTH: 969
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6245
Alignment Scores:
Pred. No.: 2,28-36 Length: 969
Score: 369.50 Matches: 91
Percent Similarity: 51.69% Conservative: 77
Best Local Similarity: 28.00% Mismatches: 102
Query Match: 13.30% Indels: 55
DB: 4 Gaps: 10

US-09-868-987-14 (1-552) x US-09-252-991A-6245 (1-969)
Qy 227 PheValGlyLeuLysHisAspPheLeuArgGlyCysLysLysLeuTrpAlaValSerGly 246
Db 76 TTATGGCATCCGCAAC-----GTGGCGTTT 102

Qy 247 SerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerVal----- 264
Db 103 GCCCTCACCCCTGATC---CTGACCGCTCATCCCTCGGGCAGCTGGTTCACCAAGGGCATC 159

Qy 265 ---LeuGlyMetAspPheLysGlyTyrrAlaPheThrPheAsnProLysGluHisGly 283
Db 160 AACTTCGGCCCTCGACTTCACCGGGGGTACGCTGATCGAGCTGACGTACGAAACAACCG--- 216

Qy 284 IleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGluAlaGlyLeu 303
Db 217 ---GCCGATCCGCAAGTGGTGGTCAATGTCT-----GGCGCC 255

Qy 304 SerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLysIleTyrrPhe 323
Db 256 GGCTACGAGGACGGCGTGGTGCAGAGCTTCGGCGACGCCCGCGAGCTGCTGGTGGGATG 315

Qy 324 SerAspLysAlaLeuSerTyrrThrLysGlnIleArgAlaSerLeuLeuLysLeuThrIle 343
Db 316 CCCAGCGAGACCGGAACTGGCAAGAGTTCGCTACCGCTTCGACGAGGCGGAT--- 372

Qy 344 MetSerTrpArgTyrrCysGlyIleValValArgAsnArgProArgPheLeuTyrrGlyAsn 363
Db 373 -----GCCGCG 378

Qy 364 SerLysArgAsnAlaLysPheTrpSerLysValSerSerLysLeuSerLysLysMetArg 383
Db 379 AACCCGGCGAACCTGAAGCGGTGCAGTACGTCCGCCCGCGAGGTGGGTGAAGAATCGCG 438

Qy 384 TyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeuTyrrValSerLeu 403
Db 439 GACCAGGCGCGCTCGGCATGCTCTCGCGTGGCGGCATCCTGCTGATCGTGGGTTTC 498

Qy 404 ArgPheGluTrpGlnTyrrAlaPheSerAlaValCysAlaLeuIleHisAspLeuAla 423
Db 499 CGCTTCAGTGAAGTTCGCCCTCGGTGGATCTCTCGCTGGTGGCAGATGCCATCATC 558

Qy 424 ThrCysAlaValLeuPheIleAlaHisPhePheLeuLysLysIleGlnIleAspLeuGln 443
Db 559 GTGATGGGCGTGTG-----TCGTTCTTC-----CAGGTGACCTTCGACCTGACC 603

```

```

Qy 444 AlaIleGlyAlaLeuMetThrValLeuGlyTyrrSerLeuAsnAsnThrLeuIleIlePhe 463
Db 604 GTCTCGCGCTGTGTGGCGGTGCTGTTACTCTGCTGACGACACCATCGTGAATCTTC 663

Qy 464 AspArgIleArgGluAsp-----AtgGlnAlaAsnLeuPheThrProMetHis 479
Db 664 GACCGGGTGGCGGAGAACTTCGCGGTGCTGCGCAAGCGCATCTGGTCGAGAACCTGAAC 723

Qy 480 ValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrAlaThr 499
Db 724 ATC-----TCCACCAAGCAGACCTTGGTACCATCGCCACCTCGGTTTCC 771

Qy 500 ThrLeuSerValLeuLeuMetLeuPheIleGlySerSerValPheAsnPheAla 519
Db 772 ACCTGCTGGCTATCGCGCGCTGCTGTTCTTCGGCGGCGACCACTGTTGCGTTTCTCC 831

Qy 520 PheIleMetThrIleGlyLeuLeuGlyThrLeuSerSerLeuTyrrIleAlaProPro 539
Db 832 ATCGCCCTGTTCGTGGGTGAATGGCGGTACCTACTCGTGCATCTACATCGCAACCTG 891

Qy 540 LeuLeuLeuPheMet 544
Db 892 GTATTGATCTGGCTG 906

RESULT 14
US-09-252-991A-6054/C
; Sequence 6054, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6054
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6054

```

```

Alignment Scores:
Pred. No.: 2,26e-36 Length: 984
Score: 369.50 Matches: 91
Percent Similarity: 51.69% Conservative: 77
Best Local Similarity: 28.00% Mismatches: 102
Query Match: 13.30% Indels: 55
DB: 4 Gaps: 10

US-09-868-987-14 (1-552) x US-09-252-991A-6054 (1-984)
Qy 227 PheValGlyLysHisAspPheLeuArgGlyCysLysLysLeuTrpAlaValSerGly 246
Db 927 TTATGGCATCCGCAAC-----GTGGCGTTT 901

Qy 247 SerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerVal----- 264
Db 900 GCGGTACCTGATC---CTGACCGCTCATCCCTCGGCGAGCTGGTTCACCAAGGGCATC 844

Qy 265 ---LeuGlyMetAspPheLysGlyTyrrAlaPheThrPheAsnProLysGluHisGly 283
Db 843 AACTTCGGCTCGACTTCACCGCGGTACGCTGATCGAGCTGACGTACGAAACAACCG--- 787

Qy 284 IleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGluAlaGlyLeu 303
Db 786 ---GCCGATCTCGCAAGTGGTGGTCAATGTCT-----GGCGCC 748

```

QY 304 SerSerArgAspPheAArgGlnThrPheGlySerSerGluLysIleLysIleTyrPhe 323  
 Db 747 GCCTACGAGGACCGCTGTGTGAGAGCTTCGGCGAGCGCCGCGAGCTGTGTGGGATG 688  
 QY 324 SerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuThrIle 343  
 Db 687 CCCAGCAGGACCGGAGCTGGGCAAGAGGTGCTACGCTTCGACGAGCGCGAT-- 631  
 QY 344 MetSerTyrArgTyrCysGlyIleValValArgAsnArgProArgPheLeuTyrClyAsn 363  
 Db 630 -----GCCGCG 625  
 QY 364 SerLysArgAsnAlaLysPheTrpSerLysValSerLysLysLeuSerLysLysMetArg 383  
 Db 624 AACCGCGGACCTGAAGCGGTGCAGTACGTCCGCCCGCAGGTGGTGAGAACTGCGC 565  
 QY 384 TyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeuTyrValSerLeu 403  
 Db 564 GACCAGGCGCGCTCGGATGCTCTGGCGGTGGCGGCATCCTGTGCTGCTGCGCTTC 505  
 QY 404 ArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAla 423  
 Db 504 CGCTTCAGTGAAGTTCGCCCTCGGTGCGATCTCTCGCTGGTGACGATGCGATCATC 445  
 QY 424 ThrCysAlaValLeuPheIleAlaHisPheLeuLysLysIleGlnIleAspLeuGln 443  
 Db 444 GTGATGGCGGTGCTG-----TCGTCTTC-----CAGGTGACCTTCGACCTGACC 400  
 QY 444 AlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAnthrLeuIleIlePhe 463  
 Db 399 GTCTCTCGCGCTGTGTGCGGTGTGTCTACTCGCTGAACGACACCATCGTATCTTC 340  
 QY 464 AspArgIleArgGluAsp-----ArgGlnAlaAsnLeuPheThrProMetHis 479  
 Db 339 GACCGGTGCGCGAGAACTTCGCGGTGCTGCGAAGCGCGATCTGTCGAGAACCTGAAC 280  
 QY 480 ValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrAlaThr 499  
 Db 279 ATC-----TCCACGACGACAGACCTGTGTGTACCATCGCCACCTCGGTTTC 232  
 QY 500 ThrLeuSerValLeuLeuMetLeuPheIleGlySerSerValPheAsnAla 519  
 Db 231 ACGTGTGTGTATCGCGCGCTGTGTCTTCGCGCGCGACACCTGTTCTGTTCTCC 172  
 QY 520 PheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProPro 539  
 Db 171 ATCCCTCTGTTCGCGGTGTAATGGCGGTACCTACTCTGCTATACATCGCCACGCTG 112  
 QY 540 LeuLeuLeuPheMet 544  
 Db 111 GTATTGATCTGGCTG 97

## RESULT 15

US-09-252-991A-6217/c  
 ; Sequence 6217, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US/09/252.991A  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 6217  
 ; LENGTH: 429  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-6217

Alignment Scores:  
 Pred. No.: 1,49e-35 Length: 429  
 Score: 358.00 Matches: 76  
 Percent Similarity: 65.43% Conservatives: 30  
 Best Local Similarity: 46.91% Mismatches: 36  
 Query Match: 12.89% Indels: 20  
 DB: 4 Gaps: 3

US-09-868-987-14 (1-552) x US-09-252-991A-6217 (1-429)

QY 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuLeuLeuTrp 100  
 Db 428 TACGCTTCTTCGGCGTGTATCGCACCGCTTGGCTGGCTTCAACATGGTATCTGGTG 369  
 QY 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120  
 Db 368 GCGTGTATGTCGATCTCTCGCGCGACCCCTGACCTGCGCGCATCGCGGTATCTGCTG 309  
 QY 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluPhe 140  
 Db 308 ACCATGGGCAATGGGTGGAGCAACGTCGTCTCTCTCGCGGATACGCGAGAACTG 249  
 QY 141 LeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160  
 Db 248 GCCAATGGCATGTCCGTCCAAAAGGCGATCCACGAGGCTTCAACCGTGCCTTACC 189  
 QY 161 IlePheAspSerAsnLeuThrValLeuAlaSerAlaLeuLeuPheLeuAspThr 180  
 Db 188 ATCTCGACGCCAACCTGACCTCGCTGTGTGGCGGCGATCTATACGCCATCGGCGACC 129  
 QY 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200  
 Db 128 GGCCCGGTGAAAGGATTTCCGCTGACCATGTCTCGCGCATCATCACCTCGATGTTT 69  
 QY 201 AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisThrGln 220  
 Db 68 GCCATCATGTGTCACCCGC----- 51  
 QY 221 LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys 240  
 Db 50 ---GCAATGGTCAACCTGATCTTCGGC---GGCGGTGACTTC-----AAGAAG 9  
 QY 241 LeuTrp 242  
 Db 8 CTGTGG 3

Search-completed: October 28, 2003, 18:09:44

Job time : 8948 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 28, 2003, 15:26:12 ; Search time 422 Seconds  
(without alignments)  
3507.948 Million cell updates/sec

Title: US-09-868-987-14

Perfect score: 2778

Sequence: 1 MVSSPILNVLKHNASVSGK.....SLYIAPPLLLFMVKENRSK 552

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1792395 seqs, 1340900451 residues

Total number of hits satisfying chosen parameters: 3584790

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+P2n.model -DEV=xlh  
-Q=/cgn2\_1/USPRO-SPool/US09868987/runat\_28102003\_152603\_23869/app\_query.fasta\_1.711  
-DB=Published Applications NA -QFMF=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09868987@cgn\_1\_1\_271\_@runat\_28102003\_152603\_23869  
-NCPu=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Length | ID   | Description                         |
|------------|-------|--------------|------|-------------------------------------|
| 1          | 2778  | 100.0        | 1864 | 9 US-09-886-468-1 Sequence 1, Appli |

|   |    |       |      |         |    |                     |                    |
|---|----|-------|------|---------|----|---------------------|--------------------|
| C | 2  | 719   | 25.9 | 1830121 | 14 | US-10-329-960-1     | Sequence 1, Appli  |
|   | 3  | 678.5 | 24.4 | 2211    | 9  | US-09-899-482-1     | Sequence 1, Appli  |
|   | 4  | 664   | 23.9 | 1584    | 10 | US-09-974-300-1840  | Sequence 1840, Ap  |
| C | 5  | 509   | 18.3 | 80557   | 15 | US-10-080-170-647   | Sequence 647, App  |
|   | 6  | 506   | 18.2 | 9025608 | 14 | US-10-156-761-1     | Sequence 1, Appli  |
| C | 7  | 429   | 15.4 | 3309400 | 10 | US-09-738-626-1     | Sequence 1, Appli  |
|   | 8  | 408.5 | 14.7 | 1848    | 14 | US-10-260-877-119   | Sequence 119, App  |
|   | 9  | 393   | 14.1 | 1671    | 10 | US-09-881-752A-95   | Sequence 95, Appl  |
|   | 10 | 350   | 12.6 | 1035    | 12 | US-09-882-227-401   | Sequence 401, App  |
|   | 11 | 342   | 12.3 | 975     | 14 | US-10-260-877-129   | Sequence 129, App  |
|   | 12 | 335   | 12.1 | 1298    | 10 | US-09-974-300-6224  | Sequence 6224, Ap  |
|   | 13 | 311.5 | 11.2 | 609     | 8  | US-08-781-986A-671  | Sequence 671, App  |
|   | 14 | 283.5 | 10.2 | 595     | 8  | US-08-781-986A-1128 | Sequence 6816, Ap  |
|   | 15 | 275.5 | 9.9  | 1104    | 14 | US-10-156-761-6816  | Sequence 6815, Ap  |
|   | 16 | 270   | 9.7  | 1752    | 14 | US-10-156-761-6815  | Sequence 1830, Ap  |
|   | 17 | 259.5 | 9.3  | 1932    | 10 | US-09-738-626-1830  | Sequence 1, Appli  |
|   | 18 | 259.5 | 9.3  | 1960    | 9  | US-09-852-053-1     | Sequence 1, Appli  |
|   | 19 | 238.5 | 8.6  | 837     | 9  | US-09-861-451A-27   | Sequence 27, Appl  |
|   | 20 | 228   | 8.2  | 1209    | 10 | US-09-738-626-1828  | Sequence 1828, Ap  |
|   | 21 | 228   | 8.2  | 1562    | 9  | US-09-852-053-2     | Sequence 2, Appli  |
|   | 22 | 177   | 6.4  | 105184  | 11 | US-09-847-513A-1    | Sequence 1, Appli  |
|   | 23 | 169.5 | 6.1  | 2691    | 10 | US-09-974-300-2170  | Sequence 2170, Ap  |
|   | 24 | 161.5 | 5.8  | 3159    | 14 | US-10-156-761-6017  | Sequence 6017, Ap  |
|   | 25 | 144   | 5.2  | 3105    | 9  | US-09-815-242-6234  | Sequence 6234, Ap  |
|   | 26 | 139.5 | 5.0  | 3168    | 9  | US-09-815-242-8091  | Sequence 8091, Ap  |
|   | 27 | 138   | 5.0  | 2190    | 10 | US-09-738-626-696   | Sequence 696, App  |
|   | 28 | 136   | 4.9  | 3099    | 9  | US-09-815-242-7048  | Sequence 7048, App |
|   | 29 | 136   | 4.9  | 1830121 | 14 | US-10-329-960-1     | Sequence 1, Appli  |
| C | 30 | 135   | 4.9  | 38734   | 11 | US-09-373-658-30    | Sequence 30, Appl  |
|   | 31 | 128.5 | 4.6  | 7888    | 12 | US-10-085-959-28    | Sequence 28, Appl  |
|   | 32 | 128   | 4.6  | 3087    | 9  | US-09-815-242-7419  | Sequence 7419, Ap  |
|   | 33 | 126.5 | 4.6  | 3123    | 9  | US-09-912-020-122   | Sequence 122, App  |
|   | 34 | 126.5 | 4.6  | 3123    | 9  | US-09-815-242-6106  | Sequence 6106, Ap  |
|   | 35 | 126.5 | 4.6  | 3123    | 12 | US-10-287-274-271   | Sequence 271, App  |
|   | 36 | 136.5 | 4.6  | 15393   | 14 | US-10-114-170-191   | Sequence 191, App  |
|   | 37 | 125.5 | 4.5  | 3078    | 12 | US-09-912-020-123   | Sequence 123, App  |
|   | 38 | 125.5 | 4.5  | 3078    | 12 | US-10-287-274-272   | Sequence 272, App  |
| C | 39 | 125.5 | 4.5  | 39061   | 12 | US-10-085-959-148   | Sequence 148, App  |
| C | 40 | 125.5 | 4.5  | 39061   | 12 | US-10-085-959-254   | Sequence 254, App  |
| C | 41 | 125.5 | 4.5  | 48254   | 12 | US-10-085-959-829   | Sequence 829, App  |
|   | 42 | 123.5 | 4.4  | 2271    | 14 | US-10-238-075-829   | Sequence 2129, Ap  |
|   | 43 | 123   | 4.4  | 3179    | 12 | US-10-156-761-2139  | Sequence 187, App  |
|   | 44 | 123   | 4.4  | 3280    | 8  | US-08-834-705-17    | Sequence 17, Appl  |
|   | 45 | 122.5 | 4.4  | 3189    | 9  | US-09-815-242-4056  | Sequence 4056, Ap  |

ALIGNMENTS

RESULT 1

US-09-886-468-1  
; Sequence 1, Application US/09886468  
; Patent No. US20020037293A1  
; GENERAL INFORMATION:  
; APPLICANT: Aventis Pasteur Limited  
; TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses there  
; FILE REFERENCE: 77813-5  
; CURRENT APPLICATION NUMBER: US/09/886,468  
; CURRENT FILING DATE: 1999-12-23  
; PRIOR APPLICATION NUMBER: 60/113,280  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,281  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,282  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,283  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,284  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,285  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/113,385  
; PRIOR FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: 60/114,050

;; PRIOR FILING DATE: 1998-12-28  
;; PRIOR APPLICATION NUMBER: 60/114,056  
;; PRIOR FILING DATE: 1998-12-28  
;; PRIOR APPLICATION NUMBER: 60/114,057  
;; PRIOR FILING DATE: 1998-12-28  
;; PRIOR APPLICATION NUMBER: 60/114,058  
;; PRIOR FILING DATE: 1998-12-28  
;; PRIOR APPLICATION NUMBER: 60/114,059  
;; PRIOR FILING DATE: 1998-12-28  
;; PRIOR APPLICATION NUMBER: 60/114,061  
;; PRIOR FILING DATE: 1998-12-28  
;; NUMBER OF SEQ ID NOS: 26  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 1  
;; LENGTH: 1864  
;; TYPE: DNA  
;; ORGANISM: Chlamydia pneumoniae  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (101)..(1756)  
US-09-868-468-1  
Alignment Scores:  
Pred. No.: 4,69e-304 Length: 1864  
Score: 2778.00 Matches: 552  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0  
US-09-868-987-14 (1-552) x US-09-868-468-1 (1-1864)  
QY 1 MetValSerSerProIleLeuAenValProLeuLysAsnHisAlaSerValSerGlyLys 20  
DB 101 ATGGTCAGCAGCCCTATTATTAACGCTCCCATTTGAAATCATGCGAGTGTCTCAGGGAAA 160  
QY 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40  
DB 161 TTATCCACCGTGAGTGAGCAAACTCGCTCAGATTTAAATCTGGAGCGATGTCCTTTT 220  
QY 41 ValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysLysGlnCysThr 60  
DB 221 GTTCCCGAGGTTCTCAGTGAAGAGACGATCTCTCTCATCTTGGAAAAAACAATGTACA 280  
QY 61 GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80  
DB 281 CAAGGCATATCTCAGCATGCTGTGGCTTGGCAATGCTTATTGTTTGTATGAGCGGTATAT 340  
QY 81 TyrArgPheGlyGlyValIleAlaSerGlyValAlaValLeuLeuAsnLeuLeuIleTrp 100  
DB 341 TATAGATTGGAGCGTCATCGCTTCGGAGCTGTTCTTCTGAATCTTTTGTCTTATCTGG 400  
QY 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120  
DB 401 GCAGCTCTACAGTATTGGATGGCCACTCACCTTGTGAGGACTCGCTGGGATTTGTTCTT 460  
QY 121 AlaMetGlyMetAlaValAlaAsnValLeuValPheGluArgIleArgGluGluPhe 140  
DB 461 GCTATGGGGATGGCCGTPAGATGCAAAATGTTCTTGTATTGAAAGAAATCCGAGAGGAATTT 520  
QY 141 LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160  
DB 521 TTATTGTCTCAAGTCTTAAATAATCTGTAGAAAAGGATATACCAAGGCTTTTGGAGCC 580  
QY 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPheLeuAspThr 180  
DB 581 ATTTTTCATTCTAACTTGACTACAGTATTGGCCTCAGCACTTCTTTCTTCTCTAGATACA 640  
QY 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200  
DB 641 GGGCCTATTAAAGGGTTGCTTTTGACATTGATTTAGGAATTTTCTCTCAATGTTTACG 700  
QY 201-AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisThrGln 220

DB 701 GCTCTTTTCATGACTAAATTTTCTTCATGCTGCTGATCAATAAGACCCCAACATACACAG 760  
QY 221 LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys 240  
DB 761 TTGCATATGATGAATAAGTTCGTGGGATAAAGCATGATTTCTTGTAGAGGATGCAAAAAA 820  
QY 241 LeuTrpAlaValSerClySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla 260  
DB 821 CTTTGGGCTGTTTCTGGAAGTGTTTTCTTTAGGTTGCTGCTCTCGGTTTGGAGCC 880  
QY 261 TrpAsnSerValLeuGlyMetAspPheLysGlyIleAlaPheThrPheAsnProLys 280  
DB 881 TGGAAATTCGTTTGGGAATGGATTTAAAGGAGGGTAAGCCTTTTACCTTTAATCCAAAA 940  
QY 281 GluHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGlu 300  
DB 941 GAGCATGGCATCAGCGATGTTGCTCAATCGCTGGCAAAAGTTGTGCATAAACTTACAGGAA 1000  
QY 301 AlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGlyLysIleLys 320  
DB 1001 GCTGGTCTTTTCTTAGAGACTTCCGTATTCAAAACATTTGGATCTTTCAGAAAAGATCAAA 1060  
QY 321 IleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLys 340  
DB 1061 ATCTATTTTAGTATGAAGCTTTAAGCTATATACTAAGCAATACGAGCCTCTCTCTTAAAA 1120  
QY 341 LeuThrIleMetSerTrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeu 360  
DB 1121 TTAACGATCATGAGTGGCGTTATTGTGGGATTTGTGTGAGAAACAGGCTTAGATTTCTC 1180  
QY 361 TyrGlyAsnSerLysArgAsnAlaLysPheTrpSerLysValSerSerLysLeuSerLys 380  
DB 1181 TAGGAAACTCTAAACGAAACGCAAAATTTGTCAAGAGGTAAAGCAGCAAACTATCGAAG 1240  
QY 381 LysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeuTyr 400  
DB 1241 AAAATGCGTTATCAGCGACCATCGGCTTTTAGGAGCTTTGGCAATCATCTTGCTCTAT 1300  
QY 401 ValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAsp 420  
DB 1301 GTGAGTTTCGGCTTGAATGGCAATATGCTTTTCAGTGGCTATGCGCTTTAATTCATGAC 1360  
QY 421 LeuLeuAlaThrCysAlaValLeuPheIleAlaHisPhePheLeuLysIleGlnIle 440  
DB 1361 CTTTGGCTTACCTGTGCTGCTGTGTTTATAGCACATTTCTTTTGAAGAAAAATTCAAATA 1420  
QY 441 AspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeu 460  
DB 1421 GATTTCGAAGCCATTTGGTCTTTTATGACTGTATTGGGCTATTCAATAACATCTTTG 1480  
QY 461 IleIlePheAspArgIleArgGluAspArgGlnAlaAsnLeuPheThrProMetHisVal 480  
DB 1481 ATCATTTTGTATCGTATTTCGTGAAGATCGCAAGCGAACCTGTTTACCCTATGCTATGTT 1540  
QY 481 LeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThrThr 500  
DB 1541 TTAGTTAATGATGCCCTTCAAAAGAGCTTCAGCCGCAAGTAAATCAACAGCTACAACT 1600  
QY 501 LeuSerValLeuLeuMetLeuLeuPheIleGlyGlySerSerValPheAsnPheAlaPhe 520  
DB 1601 CTATCAGTTTGTAAATGCTTTTGTATTAGGGGCTCTCTGCTCTTTAATTTTGCTATT 1660  
QY 521 IleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProLeu 540  
DB 1661 ATTATGACCATAGGATTTCTTCTAGGAATTTATGCTCTCTTTATATTGCACCACCTCTG 1720  
QY 541 LeuLeuPheMetValArgLysGluAsnArgSerLys 552  
DB 1721 TTGTTGTTTATGCTCGTAAGAAAAATCGCTCAAAA 1756  
RESULT 2  
US-10-329-960-1/c

```
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, .Frag
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; NAME/KEY: misc_feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (121344)..(121344)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
```





```
QY 363 AsnSerLysArgAsnAlaLysPheTrpSerLysValSerSerLysLeuSerLysLysMet 382
D 1651 TCTGATCAAAATGACACACAGTTTCACCGACAGTCGTGAAGGAGCTGGCGAGAAATGCG 1710
QY 383 ArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeuTyrValSer 402
D 1711 CTGTACGAGTGTGTATAGCTTCTATTGGC-----ATCATATTACGTTTCA 1758
QY 403 LeuArgPheGlnTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeu 422
D 1759 ATCCGATTGCAATACAAAATGCGCATTGCTGCCATCGCTCATTTGCTATATGAC----- 1812
QY 423 AlaThrCysAlaValLeuPheIleAlaHisPheLeuLysLysIleGlnIleAspLeu 442
D 1813 -----GCATTCTTTATCGTCACGCTTCTAGTATTACAAGGCTTGAGGTAGATGTT 1863
QY 443 GlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuIle 462
D 1864 ACATTATCGCGGCATCTTGACGATAATCGGTATTCCATTACGATACATCGTTTACA 1923
QY 463 PheAspArgIleArgGlu-----AspArgGlnAlaAsnLeuPheThrProMetHis 479
D 1924 TTTGACAGGTCCCGGAGCATATGAAAAAGCGTAAGCCGAAAAACCTTTGCCGATCTGAAC 1983
QY 480 ValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThr 499
D 1984 CATATTGTAACCTGACGCTGACGCAACCTTTACACGTTTCAATTACACATGTTAATACC 2043
QY 500 ThrLeuSerValLeuLeuMetLeuPheIleGlySerSerValPheAsnPheAla 519
D 2044 GTTGTGATTGTTGTGACATGCTGATCTTTGGAGCATCTTCTATTACTACTTCTCA 2103
QY 520 PheIleMetThrIleGlyLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProPro 539
D 2104 ATTGCTTTATTTGGTCGGGCTTTAAACAGCGCTTTATTCTTCTCTATACATTGCCGCACA 2163
QY 540 LeuLeuLeuPheMetValArgLysGluAsnArgSerLys 552
D 2164 ATTTGGCTTGC-ATGGAAGAAGAGAGAACTGAAAAAAGA 2201
RESULT 4
US-09-974-300-1840
; Sequence 1840, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; CURRENT APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1840
; LENGTH: 1584
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1840
```

## Alignment Scores:

|                        |          |               |      |
|------------------------|----------|---------------|------|
| Pred. No.:             | 9,52e-65 | Length:       | 1584 |
| Score:                 | 664.00   | Matches:      | 173  |
| Percent Similarity:    | 48.98%   | Conservative: | 114  |
| Best Local Similarity: | 29.52%   | Mismatches:   | 177  |
| Query Match:           | 23.90%   | Indels:       | 122  |
| DB:                    | 10       | Gaps:         | 16   |

US-09-868-987-14 (1-552) x US-09-974-300-1840 (1-1584)

```
QY 17 ValSerGlyLysPheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGly 36
D 21 ATTGAAGGCAACTTTACGATTTCAGGAAGCAAAAGACATGCGAGCATTTTAAATGCGGA 80
QY 37 AlaMetSerPheValProGluValLeuSerGluGlu-----ThrIleSerSerAsp 53
D 81 GCTCTT-----CCTGTCAAACTCGTTGAAAAGTATTCGACATCCCGTCGAGCGCAG 131
QY 54 LeuGlyLysGlnCysThrGlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeu 73
D 132 TTTGGGACGAGCTTTGAATGATACCGTGTTCGCGGAATCATTCGTATGCTATTATT 191
QY 74 IleValLeuMetSerValTyrTyrArgPheGlyValIleAlaSerGlyAlaValLeu 93
D 192 TTTCTATTATGCTTCTTACTATCGCTTGCCTCGGTTTATT-----CGGTCATT 242
QY 94 LeuAsnLeuLeuIleTrpAlaLeuGln-----TyrLeuAspAlaProLeu 110
D 243 AGCTTTTCGCTTACATCTACATTACATTCGACATTCGATTCGATGGAATCGGTTCTG 302
QY 111 ThrLeuSerGlyLeuAlaGlyIleValLeuAlaMetGlyMetAlaValAspAlaAsnVal 130
D 303 ACCCTTCGCGGATCGCGCTCTTATATTGGCGTCGGATGCGGTTTCACGCCAATC 362
QY 131 LeuValPheGluArgIleArgGluGluPheLeuLeuSerGlnSerLeuLysSerVal 150
D 363 ATTACGTATCAACGCAATAAAGAGAGCTGGAAGCTCGGCAATTCGTCGCTCGGCTTT 422
QY 151 GluLysGlyTyrThrLysAlaPheGlyAlaIlePheAspSerAsnLeuThrValLeu 170
D 423 AAAGCAGGAACAGAGATCGTTTGCACAGATTTTGTATGCCAATATTACGACGATGCTT 482
QY 171 AlaSerAlaLeuLeuPhePheLeuAspThrGlyProIleLysGlyPheAlaLeuThrLeu 190
D 483 GCGGAATCGCTTTTCATATTGTTGTAACAGCTCTGTCAAGGCTTCGCGACCATGCTG 542
QY 191 IleLeuGlyIlePheSerSerMetPheThrAlaLeuPheMetThrLysPhePheMet 210
D 543 ATCCTATCCATTTTGACAAGCTTTATTACAGCTGTCTTCCTCGCGCTTCCTGCTCGGC 602
QY 211 Leu-----TrpMetAsnLysThrGln----- 217
D 603 CTGCTGGTCGAAAGCGCTGGCTTCACCGAAAAAGAGCTGTTTGGGTGCGCAAAAAA 662
QY 218 -----HisThrGlnLeuHisMetMetAsnLysPheValGly 229
D 663 GACATTTTGGATATCAGAAAGACAGACGAAAAATACGGAGCCTCCGAAACCGTTT----- 716
QY 230 IleLysHisAspPheLeuArgGlyCysLysLysLeuTrpAlaValSerGlySerValPhe 249
D 717 TCCAAATGGGATTCGTCGGGAAGCGCAATGTTCTTGTGCAATTTTCGGCGCTTCTCTG 776
QY 250 LeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerValLeuLysMetAspPhe 269
D 777 GCGCTGGTCTTATCGTCTTGTGTTAAGCTGAAC-----CTCGCATCGATTTC 830
QY 270 LysGlyGlyTyrAlaPheThrPheAsnProLysGluHisGlyIleSerAspValAlaGln 289
D 831 TCAAGCGGA-----TCGACA 845
QY 290 MetArgGlyLysValValHisLysLeu-----GlnGlu 300
D 846 ATCGAAGTTTCAAAGCGCATATAAGCTGACACCCAGCAGCTCGAAAAAGGACTTTTGAACA 905
QY 301 AlaGlyLeu-----SerSerArgAspPheArgIleGln 311
D 906 GTCCGGCTTGATCTCGATCGTCTTTCAGCGCGCAAAAAATGATCATCGGCTCGCC 965
QY 312 ThrPhe-----GlySerSerGluLysIleLysIleTyrPheSerAsp 325
D 966 AGGTTTGTGCGGCTGCCGAATCAGAAAAAGATCGCTGAAAGTCAAGATTTATTTAAAGAC 1025
```

Qy 326 LysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeuThrIleMetSer 345  
 Db 1026 AAA-----  
 Qy 346 TrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeuTyrGlyAsnSerLys 365  
 Db 1029 -----TACGGAAGCGAACCA 1043  
 Qy 366 ArgAsnAlaLysPheTrpSerLysValSerLysLysSerLysLysMetArgTyrGln 385  
 Db 1044 AATGTAGCACCAGCTATCGCTTACCGTCGGCAAGAGCTGGCGAGAAATGCGCTTATGCG 1103  
 Qy 386 AlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeuTyrValSerLeuArgPhe 405  
 Db 1104 GTGATCATCGCTCAATCGG-----ATCATTTGTATGCTCGATCGGCTTT 1151  
 Qy 406 GluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCys 425  
 Db 1152 GAATACAAGATGGCGATTGCGCTTACTTCAATTCGTGTATGAC----- 1196  
 Qy 426 AlaValLeuPheIleAlaHisPheLeuLysLysIleGlnIleAspLeuGlnAlaIle 445  
 Db 1197 GCGTTTTTCATGTGCGCTTTCAGCCTGCAAGGCTTGAGTGTGATTTTCAATT 1256  
 Qy 446 GlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIlePheAspArg 465  
 Db 1257 GCTCCGCTGTCAGATCATCGCTATTCGATTAATGATACATTAACGTTTGACAGG 1316  
 Qy 466 IleArgGlu-----AspArgGlnAlaAsnLeuPheThrProMetHisValLeuVal 482  
 Db 1317 ATCAGGAGCATATGAAAAAGCGAAACCGAAACCTTCAGCGACTTGTGCAATATTGTG 1376  
 Qy 483 AsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThrThrLeuSer 502  
 Db 1377 AATTGAGCTTGCAGCAGACGTTACGCGTTTCGATTAATACGGTGTGACTGTGCTCAT 1436  
 Qy 503 ValLeuLeuMetLeuLeuPheIleGlySerValPheAsnPheAlaPheIleMet 522  
 Db 1437 GTCTCATGTGCATCTTCATCTTCGGCGCTGCATCGATTTCGAACCTTCTGTTGCGCTG 1496  
 Qy 523 ThrIleGlyLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProProLeuLeuLeu 542  
 Db 1497 TTGTCGACATTTTGAGCGGAGTATATTCACTCACTTACATCGCAGCTCAACTATGCTT 1556  
 Qy 543 PheMetValArgLysGlu 548  
 Db 1557 GTCTGGAAGGCGCGAA 1574

## RESULT 5

US-10-080-170-647/c  
 ; Sequence 647, Application US/10080170  
 ; Publication No. US20030129601A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: COLE, S.T.  
 ; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR  
 ; IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR  
 ; TREATMENT OF MYCOBACTERIOSES  
 ; FILE REFERENCE: 03495.0218  
 ; CURRENT APPLICATION NUMBER: US/10/080.170  
 ; CURRENT FILING DATE: 2002-06-10  
 ; PRIOR APPLICATION NUMBER: 60/270,123  
 ; PRIOR FILING DATE: 2001-02-22  
 ; NUMBER OF SEQ ID NOS: 652  
 ; SOFTWARE: PatentIn ver. 2.1  
 ; SEQ ID NO 647  
 ; LENGTH: 80557  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis  
 ; US-10-080-170-647  
 Alignment Scores:  
 Pred. No.: 2.02e-44 Length: 80557

Score: 509.00 Matches: 160  
 Percent Similarity: 42.58% Conservative: 124  
 Best Local Similarity: 23.99% Mismatches: 221  
 Query Match: 18.32% Indels: 162  
 DB: 15 Gaps: 18  
 US-09-868-987-14 (1-552) x US-10-080-170-647 (1-80557)  
 Qy 2 ValSerSerProLeuLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys--- 20  
 Db 66923 CTGACTCGCAGGTGTGTCTGACGCACGATCCAGGAACGATCCCGCGCGCGCAACC 66864  
 Qy 21 -----PheThrHisArgGluValSerLysLeuAlaSerAsp 32  
 Db 66863 CAGATCAGCGGTGGGATCCGCCATTACCCGCGCGACCCACGCCAGCTCGCAACGTC 66804  
 Qy 33 LeuLysSerGlyAlaMetSerPheValProGluValLeuSerGluGluThrIleSerSer 52  
 Db 66803 CTCGAAGTACGGGTGCTGCTCGCTGCTCTTCCGAACCGTCGGAAGCCCAACGGTTTCGCG 66744  
 Qy 53 AspLeuGlyLysLysGlnCysThrGlnGlyIleIleSerAlaCysCysGlyLeuAlaMet 72  
 Db 66743 ACCTTGGGTTGTCATCGTCGCGCGGCGATGATCGCGGTGCGATCGGCTGCTGTTG 66684  
 Qy 73 LeuIleValLeuMetSerValTyrArgPheGlyGly-----Val 86  
 Db 66683 GTGCTGTGTATTGCTGTCTTACTACCGAGTGTCTGGACTGTCTACGGCTTGTCTCGCTA 66624  
 Qy 87 IleAlaSerGlyAlaValLeuLeuLeuLeuLeuIleTrpAlaLeuGlnTyrLeu 106  
 Db 66623 GTGCTTCTCGCTCAATGTTTTCGATCTCTGCTGCTTGGGC-----CGATATATC 66570  
 Qy 107 AspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAlaMetGlyMetAlaVal 126  
 Db 66569 AACTACACCTAGATCTGCTGTGTTTTCGGGCTGATCGGTATCGGACCAACCGCC 66510  
 Qy 127 AspAlaAsnValLeuValPheGluArgIleArgGluPheLeuLeuSerGlnSerLeu 146  
 Db 66509 GACTCGTTCGTGTGTCTTCGAACGATCAAGACGAGATCCGGAAGCGCTTCTGCTTC 66450  
 Qy 147 LysLysSerValGluLysGlyTyrThrLysAlaPheGlyAlaIlePheAspSerAsnLeu 166  
 Db 66449 CGGTGCGCAGTGGCGCGTGGCGCGTCCCGCAAGACGATCGTGTGCGGCAACGCC 66390  
 Qy 167 ThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGlyProIleLysGlyPhe 186  
 Db 66389 GTCACTTCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 66330  
 Qy 187 AlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThrAlaLeuPheMetThrLys 206  
 Db 66329 GCGTTACCCCTG-----GGCTCACACGATCCTTGGACCTCGTCTGTG 66288  
 Qy 207 PhePhePheMetLeuTrp-----MetAsnLys 215  
 Db 66287 GTGTTCTTGTGACGTGGCGTGTGTATCTGCGGTCCAAATCGTCTGCTGCGCAAG 66228  
 Qy 216 ThrGlnHisThrGlnLeuHisMetMetAsnLysPheValGlyIleLys----- 231  
 Db 66227 CCGGCATACAACCGCTTGGGCGGTGCGACAGGTGCGACGCAACCGCGGCGCATGCT 66168  
 Qy 231 ----- 231  
 Db 66167 CGGACGGGCGGATAGCCGATGGCGTCCAAAGCAAGACCGCGCGGACGAGAACGCC 66108  
 Qy 231 ----- 231  
 Db 66107 ACGTGCGCGGTGGAATCACCAGGCGCACCGAGTGTCTGTGCGCGGACCGACGCGAC 66048  
 Qy 232 -----HisAspPheLeuArg----- 236  
 Db 66047 AGCACCAACAGACCGCTCGAAGCTGGGCGACCATAGCTTCTCTTCGCGCTGTACACC 65988  
 Qy 237 -----GlyCysLysLysLeuTrp---AlaValSerGlySer 247





```
Db 8161298. CGTCCCGCGGACCATCTCGTCTCCGACTTCGTGTCTCTCGCGCCGCGGTGCTG 8161357
Qy 176 PhePheLeuAspThrGlyProIleLysGlyPheAlaLeuThrLeu----- 190
Db 8161358 TTTCATCGTGACCGTCCGGAAGTCCAGGCTTCGGGTTACGGTCTGACACCGTG 8161417
Qy 191 -----IleLeuGlyIlePheSerSerMetPheThrAlaLeuPheMetThrIlys 206
Db 8161418 CTCGACGTGGTGGTCTGCTCTCTTCCACCAAGCGCTGATGACGATCTCGCCGCAAG 8161477
Qy 207 PhePhePheMet-----LeuTrpMetAsn-LysThrGlnHisThrGlnLeu----- 221
Db 8161478 CAGTTCTTCGCGAACGGTACACGTGGTCCGGCCTCGACCCGAAAGCGCTCGGCCAG 8161537
Qy 222 -----HisMetMe 224
Db 8161538 CGCCATTTCGGCGCACCGGAGTCTCTCCGCCCGCGGACCCGAGGAGGGGTGAGAT 8161597
Qy 224 tAsnLysPhe-----ValGly-----IleLysHisAspPheLe 235
Db 8161598 GTCGAAGCTCGCAACCTCGCGCCGCTCCACCGAGCGAGGTCCGCTACGACTTCGT 8161657
Qy 235 uArgGlyCysLysLysLeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAl 255
Db 8161658 C---GGCAACCGCAAGATCTGG---TACGGCATCTCGATCTCGATCACCATCACGGCCAT 8161711
Qy 255 aLeuGlyPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGlyGlyTyAlaPh 275
Db 8161712 CCTCGGCTGAGGTGGCGGCGCTGAACATGGGCATCGAGTTCAGGGCGGAGGCGTCTT 8161771
Qy 275 eThrPheAsnProLys-----GluHisGlyIleSerAspValAl 288
Db 8161772 CACC---ACCCCAAGACAGCGTCTCGTGTCCGAGCCGAGAGGTACGCGGAGGAGGC 8161828
Qy 288 aGlnMetArgGlyLysValValHisLysLeuGlnGluAlaGlyLeuSerSerArgAspPh 308
Db 8161829 GTCCGGCCACACGCGGATCGTGCAGAGCTCGGTCCGGCGGTCTCGCGTCCAGATCGC 8161888
Qy 308 eArgIleGlnThrPheGlySerSerGluLysIleLysIleTyPheSerAspLysAlaLe 328
Db 8161889 CGGTGTGACACC---GCCAAGTCGACCATCAAGCCGAGCTCTCCAGAACCTTGAA 8161945
Qy 328 uSerTyThrLysGlnIleArgAlaSerLeuLeuLysLeuThrIleMetSerTrpArgTy 348
Db 8161946 GGTGACTCGGAAGATCAACGCGGACCTGTGTGCGC-----CCGACTCG----- 8161991
Qy 348 rCysGlyIleValArgAsnArgProArgPheLeuTyArgLysSerLysArgAsnAl 368
Db 8161992 ---GGTGAGCAGATCGCAAC----- 8162009
Qy 368 aLysPheTrpSerLysValSerSerLysLeuSerLysLysMetArgTyGlnAlaThrIl 388
Db 8162010 -AAGCCCTGGCAA----- 8162021
Qy 388 eGlyLeuLeuGlyAlaLeuAlaIleIleLeuLeuTyValSerLeuArgPheGluTrpGl 408
Db 8162022 -GGCCTGTGTATCTTCATGATCCTGGTGTGATCTACCTGCGGATCGCCTTCGATGGCG 8162080
Qy 408 nTyAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLe 428
Db 8162081 GATGCGGTTCGGGCGCTTCGCGCTGATCCAGCATCACCATCAGGTTCGCGCATCTA 8162140
Qy 428 uPheIleAlaHisPhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGlyAlaLe 448
Db 8162141 CGCCCTCTCGCGCTTC-----GAGGTGAGCCCGCGGTACCGTTCATCGGTCT 8162185
Qy 448 uMetThrValLeuGlyTyTrSerLeuAsnAsnThrLeuIleIlePheAspArgIleArgGl 468
Db 8162186 GCTCAGGATCTCGGTACTCGTCTATGACACGGTGTGCTCTCGACACCCCTCAAGGA 8162245
Qy 468 u-----AspArgGlnAlaAsnLeuPheThrProMetHisValLeuValAsnAspAl 485
Db 8162246 GCAGACGAAGGACCTTCACCAAGCAGACTCGCTGGACCTTACAGCATGTGCGCAACCGCTC 8162305

Qy 485 aLeuGlnLysThrPheSerArgThrValMetThrAlaThrThrLeuSerValLeuLe 505
Db 8162306 GATCAACAGCACCTCTGGTCCGCTCCATCAACACACCGGTGGTGGTCTTCTGCGGTGCG 8162365
Qy 505 uMetLeuLeuPheIleGlyGly-----SerSerValPheAsnPheAlaPh 520
Db 8162366 GGGCTCTGCTTTCATCGCGGTGGCGTCTCGCGCGCGGACGCTCAACGACATCTCGCT 8162425
Qy 520 eIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyTrIleAlaProPtoLe 540
Db 8162426 GTGCTCTTCTGCTGGTCTCGCGCGCGGCGGTACTCTCTCATCTTTCATCGCCACGCGCT 8162485
Qy 540 uLeuLeuPheMetValArgLysGluAsnArgSerLys 552
Db 8162486 CGTCCCGCAGCTCAAGAGCGCGGAGCGCGAGATGAAG 8162522

RESULT 7
US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09738626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Alignment Scores:
Pred. No.: 9,73e-33 Length: 3309400
Score: 429.00 Matches: 142
Percent Similarity: 42.75% Conservative: 147
Best Local Similarity: 21.01% Mismatches: 222
Query Match: 15.44% Indels: 166
DB: 10 Gaps: 18

US-09-868-987-14 (1-552) x US-09-738-626-1 (1-3309400)
Qy 1 MetValSerSerProfileLeuAsn-----ValProLeuLysAsnHisAlaSerValSer 18
Db 1759690 GTGATTTCTGCACCCCGGTATTCAGTACGAAACCCCTGTGGTCTTGCAACATCCATCACC 1759631
Qy 19 GlyLysPheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAla--- 37
Db 1759630 GGTGACTTCACTCAAACTGAAGCCCAAGATCTGCGCAACAACTCGCTACCGGTGCATTG 1759571
Qy 38 ---MetSerPheValProGlu-----ValLeuSerGluGluThrIleSerSer 52
Db 1759570 CCCCTGAGCTTCGAGGTGAAGAACGCGGAGCGCGGAGAACCTACCAACCGTTCCGCCA 1759511
Qy 53 AspLeuGlyLysLysGlnCysThrGlnGlyIleIleSerAlaCysCysGlyLeuAlaMet 72
```

```
Db 1759510 TCACTAGCGCAGCATCCTTGAAGCGCGACTGATCGAGGCATCGTCCGATCGCGCTG 1759451
Qy 73 LeuileValleuMetSerValTyrArgPheGlyGlyValile----- 87
Db 1759450 GTCGCCATTCGTGTTGTCGCTACTACCGGTCTTCGGATTCGTTCCCTGTTACCCCTG 1759391
Qy 88 ---AlaSerGlyAlaValLeuLeuAsnLeuLeuLeuLeuTyrAlaAlaLeuGlnTyrLeu 106
Db 1759390 TTTGCCGAGCGGTGGTCTAGCGCTTCTGTTACTGCTGGGA-----CGCTGGATC 1759337
Qy 107 AspAlaProLeuThrLeuSerGlyLeuAlaGlyLeuValLeuAlaMetGlyMetAlaVal 126
Db 1759336 GGATATTCCTAGACCTTGTGGTATCGCGGTGTGATCATCGGTATCGGTATCCGATCCGCC 1759277
Qy 127 AspAlaAsnValLeuValPheGluArgLeuGluGluPheLeuLeuSerGlnSerLeu 146
Db 1759276 GACTCCTTCGTGGTGTCTATAGCGCATCAAGATGAGATCGGTGAAGGAAGATCCTTT 1759217
Qy 147 LysLysSerValGluLysGlyTyrThrLysAlaPheGlyAlaLeuPheAspSerAsnLeu 166
Db 1759216 AGATCTCGATCTCTGTCGTGTAAGGCGCAAGCGCCACCATCTGTCAGGCAACATG 1759157
Qy 167 ThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGlyProIleLysGlyPhe 186
Db 1759156 GTCACTTTGCTCGCGCTATCGTCAATTTACTTGTCTCGGTGCGGGAAGTCAAGGGCTTT 1759097
Qy 187 AlaLeuThrLeu---IleLeuGlyIlePheSerSerMetPheThrAlaLeuPheMet-Th 205
Db 1759096 GCCTTCACCTGGTGTGACACCGCATTCGATCTCGTCTGTGACC-----TTCTGATC 1759043
Qy 205 rLysPhePheMetLeuTrpMetAsnLysThrGlnHisThrGlnLeuHis----- 222
Db 1759042 ACGGCACCACTGGTTATCTTGGCATCAGCAACCCATCTTTGCCAAAGTATCGGTCAAC 1758983
Qy 222 -----MetMetAsnLysPheValGlyIle----- 230
Db 1758982 GGCAATGGAGAGTGATGAAGCTGTTGAAGAACCGCGCCCAACGGTGAATGGATGAG 1758923
Qy 230 -----CCTGAGTACCTGAAAAAGATCCATGCAAGAAATGCGGCACTGATAGGCTTCCACTGAC 1758863
Db 230 ----- 230
Qy 1758862 AATCTTCCACTGACATTTCTGAAGCACCTTGGACCATACGACCAAGAGGAGGAAG 1758803
Db 230 ----- 230
Qy 1758802 TAGCCATGACTGATTCCAGACTGAATCACTGTCACTCAGAGCGGTAAACCAGCCAAA 1758743
Qy 231 -----LysHis-AspPheLeuArgGlyC 238
Db 1758742 AAGCAGTGTGTTCAACAGCCTCTACACGGTGCAGCGGCATGACTTCATCCGCAAAA 1758683
Qy 238 yLysLysLeuTyrAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyP 258
Db 1758682 CCAAACTGTGTACTGATCACCAGCATTTTGTGTTATCTCGATCTCTGATCTTATCGCCA 1758623
Qy 258 heGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGlyTyrAlaPheThrPheA 278
Db 1758622 TCCGTGGCTTCTCC-----CTGAGCATCGATTTCAGGCGCGGTACCAAGATGAGCATGC 1758569
Qy 278 snProLysGluHisGlyLysSerAspValAlaGlnMetArgGlyLysValValHisLysL 298
Db 1758568 CAGCATCGGATTTACTCCACCGAAACAGGTGGAGAAACC-----TTTA 1758527
Qy 298 euGlnGluAlaGlyLeuSerSerArgAspPheArgGlnGlnThrPheGlySerSerGluL 318
Db 1758526 CTGAGCCACCGCATTTACTCCGAAATCGTGAGATCTGCGGTTCGCGGACGCCGCA 1758467
Qy 318 ystletysileTyrPhe-SerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSer 337
```

## RESULT 8

```
US-10-260-877-119
; Sequence 119, Application US/10260877
; Publication No. US20030021813A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Chovan, Linda E.
; APPLICANT: Hessler, Paul E.
; APPLICANT: Reich, Karl A.
; TITLE OF INVENTION: ESSENTIAL BACTERIA GENES AND GENOME
; TITLE OF INVENTION: SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: 'ESSENTIAL GENES'
; FILE REFERENCE: 6565.US.P1
; CURRENT APPLICATION NUMBER: US/10/260,877
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US/09/649,145
; PRIOR FILING DATE: 2000-08-25
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119
; LENGTH: 1848
; TYPE: DNA
; ORGANISM: H. influenzae
```

FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1).....(1848)  
 OTHER INFORMATION: HI-0240  
 US-10-260-877-119

Alignment Scores:  
 Pred. No.: 1,066-35 Length: 1848  
 Score: 408.50 Matches: 87  
 Percent Similarity: 65.8% Conservative: 48  
 Best Local Similarity: 42.4% Mismatches: 59  
 Query Match: 14.70% Indels: 11  
 DB: 14 Gaps: 2

US-09-868-987-14 (1-552) x US-10-260-877-119 (1-1848)

```

Qy 13 AsnHisAlaSerValSerGlyLysPhe-----Thr 22
Db 1183 AATGTTGCAACAAATTAAGGAGCTTTGGTCTTAATTTCCAAATTAAGTGGTGTGATAGC 1242
Qy 23 HisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheValPro 42
Db 1243 ATTGGGAGACACATAATCTTCTACCTATTGAATCTGGTGCAATTAATTCACCAAT 1302
Qy 43 GluValLeuSerGluGluThrIleSerSerAspLeuLysLysGlnCysThrGlnGly 62
Db 1303 CAAATTTGTTGAAGACGACCAATTCGCCCATCATTTAGGTGCGCAAAACGTAGACGAGGG 1362
Qy 63 IleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyrArg 82
Db 1363 ATTAATGCGAGCTTTTGGGGATTAGTCTGTATTGCTTTATGCTTTTACTACAAA 1422
Qy 83 PheGlyGlyValIleAlaSerGlyValLeuValLeuAsnLeuLeuLeuIleTrpAla 102
Db 1423 ATGTTGTTGATGTCAGATTTTCATGTTTATTAATGTCATTAATGTCATTAATGTCATTA 1482
Qy 103 LeuGlnTyrLeu---AspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAla 121
Db 1483 ATGCTATTATTACCGCGCGACACTTTCAATGCGGGTATTGCGGGTATCGTTTAAC 1542
Qy 122 MetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPheLeu 141
Db 1543 TTAGTATGTCAGTAGATGCAATGATTTGTAACGCTATTAAAGAAAGAAATTCGT 1602
Qy 142 LeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGlyAlaIle 161
Db 1603 AATGCTGCTTCAATTCAGACAGCCATTAATGAAGGTTTAAACGCGCATTTACTTCTATT 1662
Qy 162 PheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGly 181
Db 1663 TTTGATGCAAACTTAACCAACATCTTAACCGCAATTAATCTATACGCGGTAGGAACAGGC 1722
Qy 182 ProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThrAla 201
Db 1723 CCAATTCAAGGTTTTCGATTACGCTTTCACTTTGCTGTTGCACTTCTATGTTTACCAGC 1782
Qy 202 LeuPheMetThrLys 206
Db 1783 ATTACAGGAACCTCGC 1797

```

# RESULT 9

US-09-881-752A-95  
 Sequence 95, Application US/09881752A  
 Patent No. US20020115078A1  
 GENERAL INFORMATION:  
 APPLICANT: Kleanthous, Harold  
 APPLICANT: Al-Garawi, Amal  
 APPLICANT: Miller, Charles  
 APPLICANT: Tomb, Jean-Francois  
 APPLICANT: Oomen, Raymond P.  
 TITLE OF INVENTION: Identification of Polynucleotides  
 TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in the  
 TITLE OF INVENTION: Genome

FILE REFERENCE: 06132/041002  
 CURRENT APPLICATION NUMBER: US/09/881,752A  
 CURRENT FILING DATE: 2001-06-15  
 PRIOR APPLICATION NUMBER: US 08/833,457  
 PRIOR FILING DATE: 1997-04-01  
 NUMBER OF SEQ ID NOS: 370  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 95  
 LENGTH: 1671  
 TYPE: DNA  
 ORGANISM: Helicobacter pylori  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (50)....(1624)  
 US-09-881-752A-95

Alignment Scores:  
 Pred. No.: 5,116-34 Length: 1671  
 Score: 393.00 Matches: 95  
 Percent Similarity: 57.49% Conservative: 47  
 Best Local Similarity: 38.46% Mismatches: 99  
 Query Match: 14.15% Indels: 6  
 DB: 10 Gaps: 3

US-09-868-987-14 (1-552) x US-09-881-752A-95 (1-1671)

```

Qy 3 SerSerProIleLeuAsnValProLeu---LysAsnHisAlaSerValSerGlyLysPhe 21
Db 932 TCACCCCGGTGATTAGGAGCGTATCGTGGGGAGCGGAGGAGATTAGCGGAATTTT 991
Qy 22 ThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheVal 41
Db 992 AGCGTGCTCAAGCAGCGGATTAGCGATCCGCTTTAGGAGTGGGCGCATGAGCGCTCCC 1051
Qy 42 ProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysLysGlnCysThrGln 61
Db 1052 ATTCAGGTTTTAGAAAAAGAAATTAAGGCCCAAGTTTAGGCAAGACAGCGTTAAACT 1111
Qy 62 GlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 81
Db 1112 TCCATTATCGCTAGTTGGGGCTTTATTTAGTATGATGGCTTTATGCTGCTTTATTAC 1171
Qy 82 ArgPheGlyGlyValIleAlaSerGlyValLeuValLeuAsnLeuLeuLeuIleTrpAla 101
Db 1172 TCTATGGCGGGGTATCGCTTTGCGGTAGTGTCAATCTTTTTCATTGTGCGG 1231
Qy 102 AlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAla 121
Db 1232 GTCATGCGGATTTTGGAGCGACGCTGACTTTACCGGAATGGCGGATTTTAAACC 1291
Qy 122 MetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPheLeu 141
Db 1292 GTGGGATTCGCGGTGATGCTAATATCATCATCAACGAGCGCATTAGAGAGTCTTAAGA 1351
Qy 142 LeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGlyAlaIle 161
Db 1352 GAGATGAGGCATCGCTAAAGCGCATCCATTAGGCTATATCAATGCGAGCGGCGGAT 1411
Qy 162 PheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGly 181
Db 1412 TTTGATTTCTAATATCACTTCTTCATCGCTTTCAGTGTATTATTATACGCTTATGGCAGCA 1471
Qy 182 ProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThrAla 201
Db 1472 GCGATTAAAGGCTTTGCGCTAACTACAGCATTTGGGATTTTAGCCTCTATTATCACCGCT 1531
Qy 202 LeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisGlnLeu 221
Db 1532 ATTCTTGGCAGCGCAAGGATTTTCAAGCGCTTTTACCTAACTC-----ACTCAAACA 1585
Qy 222 HisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLysLeu 241
Db 1586 AAAAGCCTTTTACTTTTGGTTGGCGTGAATAAA-----AGAGCTTAGGAGGTTT 1636

```

QY 242 TrpAlaValSerGlySerVal 248  
 Db 1637 TGGAAATTATTCACAGCAACTA 1657

RESULT 10  
 US-09-882-227-401  
 ; Sequence 401, Application US/09882227  
 ; Publication No. US20030158396A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kleanthous, Harold  
 ; APPLICANT: Al-Garawi, Amal  
 ; APPLICANT: Miller, Charles  
 ; APPLICANT: Tomb, Jean-Francois  
 ; APPLICANT: Oomen, Raymond P.  
 ; TITLE OF INVENTION: Identification of Polynucleotides  
 ; TITLE OF INVENTION: Encoding No. US20030158396A1 Helicobacter Polypeptides in the  
 ; FILE REFERENCE: 06132/047002  
 ; CURRENT APPLICATION NUMBER: US/09/882,227  
 ; CURRENT FILING DATE: 2001-06-15  
 ; PRIOR APPLICATION NUMBER: US 08/902,615  
 ; PRIOR FILING DATE: 1997-07-29  
 ; NUMBER OF SEQ ID NOS: 638  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 401  
 ; LENGTH: 1035  
 ; TYPE: DNA  
 ; ORGANISM: Helicobacter pylori  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (25)....(993)  
 US-09-882-227-401

Alignment Scores:  
 Pred. No.: 1.74e-29 Length: 1035  
 Score: 350.00 Matches: 91  
 Percent Similarity: 50.81% Conservative: 66  
 Best Local Similarity: 29.45% Mismatches: 104  
 Query Match: 12.60% Indels: 48  
 DB: 12 Gaps: 7

US-09-868-987-14 (1-552) x US-09-882-227-401 (1-1035)

QY 244 ValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAsnSer 263  
 Db 88 GTTTCAGCAATTTCAGCGTCTAGCGTTGGGGCTTTGTTTTTCAAAGGGTTTCT--- 144  
 QY 264 ValLeuGlyMetAspPheLysGlyTyTrAlaPheThrPheAsnProLysGluHisGly 283  
 Db 145 ---TTAGGGATTGATTTCGGGGGGG----- 168  
 QY 284 lleserAspValAlaGlnMetArgGly-----LysValValHisLys 297  
 Db 169 -----AGTTGGTGTCAAGTCGCTACACTCAAAACGCCGCCCATTAAGAAGTGGCGAT 222  
 QY 298 LeuGlnGluAlaGlyLeuSerSerArgAspPheArgileGlnThrPheGlySerSerGlu 317  
 Db 223 CTGTTTGAAGAAGAGCTCGCTTCAAAGGGCTGCAAGTGGAGCGAATTTGGCTCTAAAGAA 282  
 QY 318 LysileLysileTyTrPheSerAspLysAlaLeuSerTyTrLysGlnileArgAlaSer 337  
 Db 283 GAAATTTTATCAATTCCTTTTGTAGAAACGCTGAAATGAAGATCTGAACGCT--- 339  
 QY 338 LeuLeuLysLeuThrileMetSerTrpArgTyTrCysGlyileValValArgAsnArgPro 357  
 Db 340 -----ATCGTGGGCAAC----- 351  
 QY 358 ArgPheLeuTyTrGlyAsnSerLysArgAsnAlaLysPheTrpSerLysValSerSerLys 377  
 Db 352 ---ATTCCTAAACCCAGGGGATTTTGAATCCGTAATTTGACACCGGTGGGCGCTAGA 408  
 QY 378 LeuSerLysLysMetArgTyTrGlnAlaThrileGlyLeuLeuGlyAlaLeuAlailelle 397

Db 409 GTGGGGAGCGCAATTGAAAGAGAAAGCAATTTTTCGCTGATTTTAGCATTAATAGCGATC 468  
 QY 398 LeuLeuTyTrValSerLeuArgPheGluTrpGlnTyTrAlaPheSerAlaValCysAlaLeu 417  
 Db 469 ATGGTTTATGTGAGTTTCCGCTATGAATGCGGTTTTCGCTTACGAGCGTCATTGCGCTT 528  
 QY 418 lIeHisAsp-----LeuLeuAlaThrCysAlaValLeuPheileAlaHisPheLeu 435  
 Db 529 GTGATGATGTGATTTTAGTGGCAAGCTCGGTGATTTT----- 570  
 QY 436 LysLysileGlnileAspLeuGlnAlaileGlyAlaLeuMetThrValLeuGlyTyTrSer 455  
 Db 571 ---AAGATGATATGAATTTGGAAGTGATTGCGGCTTGTCTACCTTGATTGGGTATTC 627  
 QY 456 LeuAsnAsnThrleullellePheAspArgileArgGluAspArgGlnAlaAsnLeuPhe 475  
 Db 628 ATTAATGATACGATCATTTATTTTGACAGATCAGAGAGAGATGCTTCTCAAAAAACC 687  
 QY 476 ThrProMetHisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMet 495  
 Db 688 AAAACGCCCACTCAAGCCATTGATGAAGCCATTCTTACAGCTCAGCGCAGCTTTTA 747  
 QY 496 ThrThrAlaThrThrLeuSerValLeuLeuMetLeuLeuPheileGlyGlySerSerVal 515  
 Db 748 ACTTCTTTAACCGTGTTTTGTGGTGTGATTTCGTCGTTGGAGTAAGATCATC 807  
 QY 516 PheAsnPheAlaPheileMetThrileGlyileLeuLeuGlyThrLeuSerSerLeuTyTr 535  
 Db 808 ATTGGCTTTTCATTGCCCATTGTAATAGCAGCAATTGTAGGGACTTATAGCTCTATTTTC 867  
 QY 536 lIeAlaProProLeuLeuLeuPheMet 544  
 Db 868 ATCGCCCTAAAGTGGCTTATTGTTA 894

RESULT 11  
 US-10-260-877-129  
 ; Sequence 129, Application US/10260877  
 ; Publication No. US20030021813A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Abbott Laboratories  
 ; APPLICANT: Chovan, Linda E.  
 ; APPLICANT: Hessler, Paul E.  
 ; APPLICANT: Reich, Karl A.  
 ; TITLE OF INVENTION: ESSENTIAL BACTERIA GENES AND GENOME  
 ; TITLE OF INVENTION: SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF  
 ; FILE REFERENCE: 6565.US.P1  
 ; CURRENT APPLICATION NUMBER: US/10/260,877  
 ; CURRENT FILING DATE: 2002-09-30  
 ; PRIOR APPLICATION NUMBER: US/09/649,145  
 ; PRIOR FILING DATE: 2000-08-25  
 ; NUMBER OF SEQ ID NOS: 137  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 129  
 ; LENGTH: 975  
 ; TYPE: DNA  
 ; ORGANISM: H. influenzae  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)....(975)  
 ; OTHER INFORMATION: HI-0239  
 US-10-260-877-129

Alignment Scores:  
 Pred. No.: 1.27e-28 Length: 975  
 Score: 342.00 Matches: 98  
 Percent Similarity: 51.90% Conservative: 80  
 Best Local Similarity: 28.57% Mismatches: 111  
 Query Match: 12.31% Indels: 54  
 DB: 14 Gaps: 15

US-09-868-987-14 (1-552) x US-10-260-877-129 (1-975)



Qy 201 AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisThrCln 220  
 Db 1195 CGTGTCTATGGTTCAGGGTGTGCTAGCCCTCTGGGTAACAGC----- 1239  
 Qy 221 LeuHisMetMetAsnLys-----PheValGlyIleLys 231  
 Db 1240 ---AAATTTTAAATAAACGGCTGGCTGGTGTGGCGTGAAA 1278

## RESULT 13

US-08-781-986A-671  
 ; Sequence 671, Application US/08781986A  
 ; Publication No. US20030054436A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Charles Kunsch  
 ; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
 ; NUMBER OF SEQUENCES: 5255  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/781,986A  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Benson, Bob  
 ; REGISTRATION NUMBER: 30,446  
 ; REFERENCE/DOCKET NUMBER: PB248PP  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 671:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 609 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; US-08-781-986A-671

## Alignment Scores:

| Pred. No.:             | Length: | Score: |
|------------------------|---------|--------|
| 1.69e-25               | 609     | 311.50 |
| Matches:               | 62      |        |
| Percent Similarity:    | 64.57%  |        |
| Best Local Similarity: | 35.43%  |        |
| Query Match:           | 11.21%  |        |
| DB:                    | 8       |        |

US-09-868-987-14 (1-552) x US-08-781-986A-671 (1-609)

Qy 378 LeuSerLysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIlelle 397  
 Db 8 ATTCGACAGGAAATAGCAAAAATGCAATGCGATTAATCTATGATGATGATGATC 67  
 Qy 398 LeuLeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeu 417  
 Db 68 ATCATCTATGATCATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 127  
 Qy 418 IleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPheLeuLysLys 437  
 Db 128 TTATCATGATGATTC-----ATTATAGTAGCGATTTTTCAGTTTATTATTAGA 172

Alignment Scores:

| Pred. No.:             | Length: | Score: |
|------------------------|---------|--------|
| 2.41e-22               | 595     | 293.50 |
| Matches:               | 62      |        |
| Percent Similarity:    | 56.90%  |        |
| Best Local Similarity: | 35.63%  |        |
| Query Match:           | 10.21%  |        |
| DB:                    | 8       |        |

Qy 438 IleGlnIleAspLeuGlnAlaIlelleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsn 457  
 Db 173 ATTGAAGTAGATTAAACATTTATCGCGCTGTATTACAAATTTGCGGTATTCAATTAAT 232  
 Qy 458 AsnThrLeuIlelePheAspArgIleArgGluAspArgGln---AlaAsnLeuPheThr 476  
 Db 233 GATACAATCGTAACGTTTACCGGTGACGTGAAACCTTCAAAAGGTTAAAGTGATTACG 292  
 Qy 477 -----ProMethisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThr 493  
 Db 293 ACAACAGAACAAATTCATGATATATCGTTAATAGATCAATTAGACAGACAATGACACGTTCA 352  
 Qy 494 ValMetThrThrAlaThrLeuSerValLeuMetLeuLeuPheIleGlyGlySer 513  
 Db 353 ATTAATACAGTATTACAGTTATTGTAGTAGTAGTATCTATCTATCTCTCGGTGCTCCT 412  
 Qy 514 SerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSer 533  
 Db 413 ACGATATTCAACTTTACTTTTAGCATTTATTTATCGGATTGATTCTCGGTGATTCTCTTCA 472  
 Qy 534 LeuTyrIleAlaProLeuLeuLeuPheMetValArgLysGlu 548  
 Db 473 ATCTTCATTGCGCTTCGCTATGGGAATAATGAAAAACGTCAG 517

## RESULT 14

US-08-781-986A-1128  
 ; Sequence 1128, Application US/08781986A  
 ; Publication No. US20030054436A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Charles Kunsch  
 ; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
 ; NUMBER OF SEQUENCES: 5255  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/781,986A  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Benson, Bob  
 ; REGISTRATION NUMBER: 30,446  
 ; REFERENCE/DOCKET NUMBER: PB248PP  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 1128:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 595 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; US-08-781-986A-1128

## Alignment Scores:

| Pred. No.:             | Length: | Score: |
|------------------------|---------|--------|
| 2.41e-22               | 595     | 293.50 |
| Matches:               | 62      |        |
| Percent Similarity:    | 56.90%  |        |
| Best Local Similarity: | 35.63%  |        |
| Query Match:           | 10.21%  |        |
| DB:                    | 8       |        |

Best Local Similarity: 25.44% Mismatches: 125  
Query Match: 9.92% Indels: 59  
DB: 14 Gaps: 12

US-09-868-987-14 (1-552) x US-10-156-761-6816 (1-1104)

|    |     |  |     |
|----|-----|--|-----|
| Qy | 230 | IleuysHisaspheLeuArgGlyCysuysuysLeuTrpAlaValSerGlySerValPhe    | 249 |
| Db | 46  | GTCCGCTACGACTTCGTC--GGCAACCGCAAGACTCG--TACGGCATCTCGATCTCGT     | 99  |
| Qy | 250 | LeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAenSerValLeuGlyMetAspPhe   | 269 |
| Db | 100 | ATCACCATCAGGCCATCTCCGGCTGACGGTCGGCGCTGAACATGGCATCGAGTTC        | 159 |
| Qy | 270 | LysGlyGlyTyrAlaPheThrPheAsnProLys-----GluHis                   | 282 |
| Db | 160 | CAGGGCGGCGGTCTTCACC--ACCCCAAGACACCGCTCCGTCGTCGAGCCGAG          | 216 |
| Qy | 283 | GlyIleSerAspValAlaGlnMetArgGlyuysValValHisLeuGlnGluAlaGly      | 302 |
| Db | 217 | AGGTACGCGGAGGAGCGCTCCGCCACGACGCGATCGTCAGAAGCTCGGTCGGTCGGCGGT   | 276 |
| Qy | 303 | LeuSerSerArgaspPheArgIleGlnThrPheGlySerSerGluLysIleuysIleTyr   | 322 |
| Db | 277 | CTCGCGCTGCAGATCGCGGTGTGGACACC--GGCAAGTCCGACGATCAAGGCCGAG       | 333 |
| Qy | 323 | PheserAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeuThr   | 342 |
| Db | 334 | CTCTCCAGAACCTGAAGTCGACTCGAGNAGATCAACGCCGACCTCGTCGGC-----       | 387 |
| Qy | 343 | IleMetSerTrpArgTyrCysGlyIleValValArgAenArgProArgPheLeuTyrGly   | 362 |
| Db | 388 | ---CCCAGCTGG-----GGTGAGCAGATCGCCAAC-----                       | 414 |
| Qy | 363 | AsnSerLysArgAenAlaLysPheTrpSerLysValSerSerLysLeuSerLysLysMet   | 382 |
| Db | 415 | -----AAGGCGCTGCAA-----   | 426 |
| Qy | 383 | ArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeuTyrValSer      | 402 |
| Db | 427 | -----GGCTGGTGATCTTCATGATCTTCGTCGTGATCTTACCTGGCG                | 468 |
| Qy | 403 | LeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeu   | 422 |
| Db | 469 | ATCGCTTCGAGTGGCGGATGGCGGTCCGGCCCTGGTCGCGCTGATCCACGACATCACC     | 528 |
| Qy | 423 | AlaThrCysAlaValLeuPheIleAlaHisPhePheLeuLysLysIleGlnIleAspLeu   | 442 |
| Db | 529 | ATCACGCTGGCATCTACGCCCTCGTCGGCTTC-----GAGTGACGCC                | 573 |
| Qy | 443 | GlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAenThrLeuIleIle   | 462 |
| Db | 574 | GGTACGGTCACTCGGTCTGCTCACGATCCTCGGTTACTCGCTCTATGACACGGTCTCGTC   | 633 |
| Qy | 463 | PhesPArgIleArgGlu-----AspArgGlnAlaAenLeuPheThrProMetHis        | 479 |
| Db | 634 | TTCCAGACCTCAAGGAGCAGACGACGACCTCCACCAAGCAGACTCGCTGGACCTACACC    | 693 |
| Qy | 480 | ValLeuValAenAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThr   | 499 |
| Db | 694 | GATGTGCCCAACCGCTCGATCAACAGCACCTCTGGTCGGCTCCATCAACACCAACGGTGGTC | 753 |
| Qy | 500 | ThrLeuSerValLeuLeuMetLeuLeuPheIleGlyGly-----SerSer             | 514 |
| Db | 754 | GCCTTCTCGCGGTCCGGCGCTCTGCTTCATCGCGGTGGCGTCTCGCGCGCGGACAG       | 813 |
| Qy | 515 | ValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeu   | 534 |
| Db | 814 | CTCAACGACATCTCGCTGTCTCGCTGTTCTCGGCTCTCGCGCGCGGCGGTACTCTCTCGATC | 873 |
| Qy | 535 | TyrIleAlaProProLeuLeuLeuPheMetValArgLysGluAsnArgSerLys         | 552 |
| Db | 874 | TTCATGCCACGCGCTCTGTCGCGCGACTCAAGGAGGCGGACGCGCATGAAG            | 927 |

Search completed: October 28, 2003, 18:57:46  
Job time : 12261 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 28, 2003, 15:26:12 ; Search time 3261 Seconds  
(without alignments)  
4114.097 Million cell updates/sec

Title: US-09-868-987-14  
Perfect score: 2778  
Sequence: 1 MVSSPILNVLKHNHASVSGK.....SLYIAPLLLPWVRKRNRSK 552

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -MODEL=frame+p2n model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spo3/US09868987/runat\_28102003\_152602\_23844/app\_query.fasta\_1.711  
-DB=EST -OPWT=fastap -SUFFIX=first -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09868987 @CNC 1.1.2810 runat\_28102003\_152602\_23844 -NCPU=6 -ICPU=3  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_mam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rtd.\*  
26: em\_gss\_phg.\*  
27: em\_gss\_vrl.\*  
28: gb\_ges1.\*

29: gb\_gse2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description         |
|------------|-------|-------------|--------|-------|---------------------|
| C 1        | 404   | 14.5        | 628    | 28    | AY080093            |
| C 2        | 388.5 | 14.0        | 934    | 29    | AY080093            |
| C 3        | 275.5 | 9.9         | 1189   | 29    | CC128762 ND.L.80M15 |
| C 4        | 244   | 8.8         | 654    | 14    | BZ560638 pac82-164  |
| C 5        | 230.5 | 8.3         | 731    | 28    | CB483547 cclwbb10   |
| C 6        | 221   | 8.0         | 627    | 29    | BH397261 AG-ND-147  |
| C 7        | 213   | 7.7         | 661    | 29    | BZ564006 pac82-164  |
| C 8        | 203.5 | 7.3         | 602    | 28    | BZ564007 pac82-164  |
| C 9        | 203.5 | 7.3         | 602    | 29    | BH182894 022 B.07-  |
| C 10       | 191.5 | 6.9         | 1124   | 29    | AL619845 T3 end of  |
| C 11       | 182   | 6.6         | 1348   | 29    | BZ560773 pac82-164  |
| C 12       | 168   | 6.0         | 548    | 28    | BZ580082 msh2-928.  |
| C 13       | 158.5 | 5.7         | 662    | 14    | AZ303132 GSSBfr0207 |
| C 14       | 151.5 | 5.5         | 1101   | 29    | CB852238 UI-CF-FNO  |
| C 15       | 150.5 | 5.4         | 655    | 28    | AL052099 Drosophila |
| C 16       | 150   | 5.4         | 1514   | 29    | BH172283 AG-ND-101  |
| C 17       | 149   | 5.4         | 1305   | 29    | BZ575576 msh2-4527  |
| C 18       | 135.5 | 4.9         | 1587   | 29    | BZ554487 pac81-60   |
| C 19       | 126   | 4.5         | 329    | 9     | BZ569745 pac82-164  |
| C 20       | 122   | 4.4         | 297    | 9     | AW064237 SP0690 KR  |
| C 21       | 122   | 4.4         | 559    | 14    | AW064146 SP0380 KR  |
| C 22       | 120   | 4.3         | 855    | 14    | CA548282 C0803G06-  |
| C 23       | 120   | 4.3         | 924    | 13    | CA985664 AGENCOURT  |
| C 24       | 120   | 4.3         | 943    | 14    | BU908977 AGENCOURT  |
| C 25       | 120   | 4.3         | 956    | 13    | CA985678 AGENCOURT  |
| C 26       | 120   | 4.3         | 1418   | 29    | BQ917610 AGENCOURT  |
| C 27       | 120   | 4.3         | 3431   | 11    | BZ573318 msh2 305.  |
| C 28       | 119   | 4.3         | 187    | 28    | BZ573318 msh2 305.  |
| C 29       | 114.5 | 4.1         | 537    | 14    | AK083566 Mus muscu  |
| C 30       | 113.5 | 4.1         | 898    | 29    | BH371348 AG-ND-132  |
| C 31       | 113   | 4.1         | 707    | 14    | CB393013 OSTR111D1  |
| C 32       | 113   | 4.1         | 3459   | 11    | CC120556 ND.L.73K8. |
| C 33       | 112   | 4.0         | 980    | 14    | CA748726 UI-H-F71-  |
| C 34       | 111   | 4.0         | 1813   | 12    | BC023740 Mus muscu  |
| C 35       | 110.5 | 4.0         | 1201   | 13    | CB206092 AGENCOURT  |
| C 36       | 109.5 | 3.9         | 672    | 29    | BG843641 1024003E1  |
| C 37       | 109.5 | 3.9         | 750    | 28    | BA422778 BX422778   |
| C 38       | 109   | 3.9         | 818    | 14    | AJ227358 Ciona int  |
| C 39       | 108.5 | 3.9         | 2486   | 29    | BH250389 BOGAB19TR  |
| C 40       | 107.5 | 3.9         | 2080   | 11    | CA432012 UI-H-FG1-  |
| C 41       | 107   | 3.9         | 676    | 28    | BZ558678 pa98401.3  |
| C 42       | 106   | 3.8         | 1031   | 12    | AY105332 Zea mays   |
| C 43       | 105.5 | 3.8         | 1331   | 10    | BH898492 MB618A6    |
| C 44       | 105.5 | 3.8         | 1662   | 10    | BM557488 AGENCOURT  |
| C 45       | 105   | 3.8         | 899    | 13    | BG326851 602425735  |
|            |       |             |        |       | BG114690 602315331  |
|            |       |             |        |       | BU159291 AGENCOURT  |

# ALIGNMENTS

RESULT 1  
AY080093/c  
LOCUS AY080093  
DEFINITION AY080093 Scripps Pier (La Jolla, CA) uncultured virus community  
sequence.  
ACCESSION AY080093  
VERSION AY080093.1 GI:24745274  
KEYWORDS GSS.  
SOURCE uncultured marine virus  
ORGANISM uncultured marine virus  
REFERENCE 1 (bases 1 to 628)  
AY080093 628 bp DNA linear GSS 06-NOV-2002  
uncultured marine virus genomic clone SI05IP3C7L, genomic survey



```

Db      756 GGC---GGTTTCATCGCCCAAGGCGCTGAACCTCGGTATCGATCTCCAGGGCGGCTCCGTC 700
Qy      275 PhetrPheAnPProLysGluHISglYleSerAspValAlaGlnMetArgGlyLysVal 294
Db      699 ATCGAAGTCCGCGCAACAG---GGTGAGCCGATCTGGCCGATATCCGCGATCGCTA 643
Qy      295 ValHisLysLeuGlnGluAlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGly 314
Db      642 AACGAG-----CTAACTGGGTGAATCCAGGCTCAGAATTTCGGT 601
Qy      315 SerSerGluLysIleLysIleTyrPhe-----SerAspLysAlaLeu 328
Db      600 ACGCGCAGGATGTTCTTATCCGATTCAGGCGCAGGAGGCGCGAAATCCGAGCAA 541
Qy      329 SerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeuThrIleMetSerTrpArgTyr 348
Db      540 TCCGCCATCACATCGTCGGCGGAGCTT----- 511
Qy      349 CysGlyIleValValArgAsnArgProArgPheLeuTyrGlyAsnSerLysArgAsnAla 368
Db      510 -----GAGGCAATATGACTTC 493
Qy      369 LysPheTrpSerLysValSerSerLysLeuSerLysLysMetArgTyrGlnAlaThrIle 388
Db      492 CGCGCGGTGAAGTTGTCCGCCCGCCGCTTCCGGTGACCTCACCTTCACCTCCACGATC 433
Qy      389 GlyLeuLeuGlyAlaLeuAlaIleLeuLeuTyrValSerLeuArgPheGluTrpGln 408
Db      432 GGCATCTTCGTCGGCATACGCGGATCATGATCTACATCTGGGTTCGCTTCGAATGGCAG 373
Qy      409 TyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeu 428
Db      372 TTTCGGCTCGCGCGCTATTTCGATGGTGACGAGTGGTCTTCACA----- 325
Qy      429 PheIleAlaHisPhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGlyAlaLeu 448
Db      324 --ATCGCCCTCTTCGTCTTCCTGGGATAGAGTTCAACCTTCACGATGTCGGCGGATC 268
Qy      449 MetThrValLeuGlyTyrSerLeuAsnThrLeuIleIlePheAspArgIleArgGlu 468
Db      267 CTGACGATTATCGGTATTCGCTGAATGACACCGCTGCATCTATGACCGTATCCGGAA 208
Qy      469 AspArgGlnAlaAsnLeuPheThrProMetHisValLeuValAsnAspAlaLeuGlnLys 488
Db      207 AACCTCAGCGCTATAGAAGATCGCGTGTGATGATATTGAGTGTCTCTGAACGAC 148
Qy      489 ThrPheSerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMetLeuLeu 508
Db      147 ACGCTGTCCGCGCACCATCTCACCGGTTGACCGTGTCTTCGGCGCTCTCGCTCTAC 88
Qy      509 PheIleGlyGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeu 528
Db      87 CTCCTCGGTGCGAGGTACGCTCTTCACCTTCGCCCATGCTCTTCGGTGTGCGGATC 28
Qy      529 GlyThrLeuSerSerLeuTyrIle 536
Db      27 GGTGTGTCTCGTCTCTACATT 4

```

```

RESULT 3
BZ560638 1189 bp DNA linear GSS 17-DEC-2002
LOCUS pacs2-164_2652.xl pacs2-164 Pseudomonas aeruginosa genomic clone
DEFINITION pacs2-164_2652, Genomic survey sequence.
ACCESSION BZ560638
VERSION BZ560638.1 GI:27179032
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1189)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,

```

```

TITLE Burns,J.L., Kaul,R. and Olsen,M.V.
JOURNAL Whole-Genome-Sequence Variation among multiple isolates of
COMMENT Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 20622216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
source
1. 1189
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone_lib="pacs2-164_2652"
/notes="clinical isolate 2-164 Whole genomic shotgun
library."
BASE COUNT 220 a 361 c 362 g 244 t 2 others
ORIGIN

```

```

Alignment Scores:
Pred. No.: 1,27e-19 Length: 1189
Score: 275.50 Matches: 54
Percent Similarity: 64.19% Conservative: 41
Best Local Similarity: 36.49% Mismatches: 40
Query Match: 9.92% Indels: 13
DB: 29 Gaps: 4
US-09-868-987-14 (1-552) x BZ560638 (1-1189)

```

```

Qy      401 ValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAsp 420
Db      43 GTCCGCTCCGCTTCCAGTGGAGTTCGCCCTCGGTGCGATCTCTCGTGGTGCACGAT 102
Qy      421 LeuLeuAlaThrCysAlaValLeuPheIleAlaHisPheLeuLysLysIleGlnIle 440
Db      103 GCGATCATCGTGTGCGGTGCTG-----TCGTTCTTC-----CAGGTGACCTTC 147
Qy      441 AspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnThrLeu 460
Db      148 GACTGACCGCTCTCGCGCTGTGCTGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 207
Qy      461 IleIlePheAspArgIleArgGluAsp-----ArgGlnAlaAsnLeuPheThr 476
Db      208 GTGATCTTCGACCGGTGCGCGAGAACTTCGCGGTGCTGCGCAAGCGCGATCTGGTCGAG 267
Qy      477 ProMetHisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThr 496
Db      268 AACCTGAACATC-----TCCACGACGACAGACCTGTTGCTGCTATATCGGCACC 315
Qy      497 ThrAlaThrThrLeuSerValLeuLeuMetLeuLeuPheIleGlyGlySerSerValPhe 516
Db      316 TCGGTTTCACGCTGCTGGCTATCGCGCGCTGCTGCTTCGCGCGCGACACCTGCTC 375
Qy      517 AsnPheAlaPheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIle 536
Db      376 GGTTTCTCCATCCGCTGTTGCTGGGTGTAATGGCGGTACTTACTCTGCTGCTATCATC 435
Qy      537 AlaProProLeuLeuLeuPheMet 544
Db      436 GCCAACGTGGTATTGATCTGGCTG 459

```

```

RESULT 4
CB483547/c 654 bp mRNA linear EST 01-APR-2003
LOCUS cclwbtbl008006 Coregonus clupeaformis head Coregonus clupeaformis
DEFINITION cDNA, mRNA sequence.
ACCESSION CB483547

```



```

DB:                28                Gaps:                9
US-09-868-987-14 (1-552) x BH397261 (1-731)

QY 114 GlyLeuAlaGlyIleValLeuAlaMetGlyMetAlaValAspAlaAsnValLeuValPhe 133
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 727 GGTATCGGGGTATCGTATTATTCATGCCAAGGAGTAGATACGAACGTATCATTTAT 668

QY 134 GluArgIleArgGluGluPheLeuLeuSerGlnSerLeuLeuLysSerValGluLysGly 153
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 667 GAAAGACCAAGAGAGATTATTCGAGGAAAAGAAATCCGTAAGCAGCATACACCCAGGT 608

QY 154 TyrThrLysAlaPheGlyAlaIlePheAspSerAsnLeuThr-ThrValLeuAlaSerAl 173
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 607 TTCAAGCAGCGTTTATCTCGATTATTGACGCACACGCCACACATTACTACCAGCTG 548

QY 173 aleuLeuPhePheLeuAspThrGlyProIleLysGlyPheAla-LeuThrLeuLeuLeuG 193
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 547 TTGTATTATATTTTCGGTACAGGACCAATCCAAAGGATTTGCCGGAACCTTAATTATCG 488

QY 193 lYilePheSerSerMetPheThrAlaLeuPheMetThrLysPhePheMetLeuTrpM 213
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 487 GTATTTTGATGACGTTCTTACTTCTGTATTATGTCGAGAGTAATGATCTTCACGAC 428

QY 213 etAsnLysThrGlnHisThrGlnLeu-----HisMetMetAsnLysPheValGlyI 230
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 427 TTGGAAGAGGTAAGAGATTTTCAGTATGGACATCTTTCTCAAAAATCTTTTCAGAAATA 368

QY 230 leLysHisAspPheLeuArgGlyCysLysLysLeuTrpAla---ValSerGlySerValP 249
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 367 TCTGGATCGACTTTATCGGG-----AAAAGAAAATGGTCTTACATCTTCTACTATAT 314

QY 249 heLeuLeuGlyCysValAla-----LeuGlyPheGlyAlaIleTrpAsnSerValL 265
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 313 TGATGATTATCTGTATCGCATCTATTGTTACAAAAGGCTC-----AAAT 269

QY 265 euGlyMetAspPheLysGlyGly-----TyrAlaPheThrPheAsnProLysGluHisG 283
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 268 TTGGGGTTGACITCAAGAGGAGGAAGTATGTTGTAAGATTTCAT---AAGCCGGTTG 212

QY 283 lYleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnAlaGlyL 303
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 211 TGGCTTCGGATATTCAGGAAGAGTTGGCTCCGATAATA---AAACTAAT-GACGCTAATA 156

QY 303 euSerSerArgAspPheArgIleGlnThrPheGlySerSerCluLysIleLysIle 321
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 155 ACGAAGCTGTAGAT-----GTTAAACATTTGGTAACTCTTAATCAGCTTAGAATA 106

RESULT 6
BZ564006          627 bp      DNA      linear      GSS 17-DEC-2002
LOCUS            pacs2-164_4490.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
DEFINITION       BZ564006
ACCESSION        BZ564006
VERSION          BZ564006.1 GI:27188954
KEYWORDS         GSS.
SOURCE           Pseudomonas aeruginosa
ORGANISM         Pseudomonas aeruginosa
REFERENCE        1 (bases 1 to 627)
AUTHORS          Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
                  Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE            Whole-Genome-Sequence variation among multiple isolates of
                  Pseudomonas aeruginosa library
JOURNAL          J. Bacteriol., (2002) In press
COMMENT          Contact: Chris K. Raymond
                  Genome Center
                  University of Washington
                  Box 352145, Seattle, WA 98105-2145, USA
                  Tel: 2062216954
                  Fax: 2066857244
                  Email: craymond@u.washington.edu

Class: shotgun.
Location/Qualifiers
1..627
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone_lib="pacs2-164_4490"
/clone_lib="pacs2-164"
/notes="clinical isolate 2-164 whole genomic shotgun
library."

BASE COUNT      100 a      196 c      182 g      149 t
ORIGIN
Alignment Scores:
Pred. No.:      6.3e-14      Length:      627
Score:          221.00      Matches:    44
Percent Similarity: 65.57%      Conservative: 36
Best Local Similarity: 36.07%      Mismatches: 32
Query Match:    7.96%      Indels:     10
DB:             29      Gaps:      3

US-09-868-987-14 (1-552) x BZ564006 (1-627)
QY 427 ValLeuPheIleAlaHisPhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGly 446
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 109 GTGATGGGGGTGCTGCTCTTCTC-----CAGGTGACCTTCGACCTGACCGTCTCGCC 162

QY 447 AlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIlePheAspArgIle 466
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 163 GCTGTGCTGGCGGTGCTGCTTACTCGGTGAACACACATCGTGTCTTCGACCGGGTG 222

QY 467 ArgGluAsp-----ArgGlnAlaAsnLeuPheThrPrometHisValLeuVal 482
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 223 CGCAGAACTTCCCGCTGCTCGCAAGCCGATCTGTCGAGAACCTTGAACATC----- 276

QY 483 AsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThrThrLeuSer 502
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 277 -----TCCACACCCAGACCTGTTGGTACTATCGCACCTCGGTTTCACCGTGTG 330

QY 503 ValLeuLeuMetLeuLeuPheIleGlyGlySerSerValPheAsnPheAlaPheIleMet 522
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 331 GCTATCGCGCGGTGCTGTTCTTCGGGGGACAACTGTTGGTTCCTCATCGCCCTG 390

QY 523 ThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProProLeuLeuLeu 542
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 391 TTCGTGCGTGAATGCGGGGTACCTACTCGTCGATCTACATCGCCCAACGTCGTATTGATC 450

QY 543 PheMet 544
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 451 TGGCTG 456

RESULT 7
BZ564007          661 bp      DNA      linear      GSS 17-DEC-2002
LOCUS            pacs2-164_4490.x3 pacs2-164 Pseudomonas aeruginosa genomic clone
DEFINITION       BZ564007
ACCESSION        BZ564007
VERSION          BZ564007.1 GI:27188956
KEYWORDS         GSS.
SOURCE           Pseudomonas aeruginosa
ORGANISM         Pseudomonas aeruginosa
REFERENCE        1 (bases 1 to 661)
AUTHORS          Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
                  Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE            Whole-Genome-Sequence variation among multiple isolates of
                  Pseudomonas aeruginosa library
JOURNAL          J. Bacteriol., (2002) In press
COMMENT          Contact: Chris K. Raymond
                  Genome Center
                  University of Washington

```

Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun

# FEATURES

source  
 1. .661  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="2-164"  
 /db\_xref="taxon:287"  
 /clone="pacs2-164\_4490"  
 /clone\_lib="pacs2-164"  
 /note="clinical isolate 2-164 Whole genomic shotgun library"

BASE COUNT 105 a 209 c 184 g 163 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 5,3e-13 Length: 661  
 Score: 213.00 Matches: 43  
 Percent Similarity: 63.93% Conservative: 35  
 Best Local Similarity: 35.25% Mismatches: 34  
 Query Match: 7.67% Indels: 10  
 DB: 29 Gaps: 3

US-09-868-987-14 (1-552) x BZ564007 (1-661)

Qy 427 ValLeuPheLeuAlaHisPhePheLeuLysLysLeuGlnIleAspLeuGluAlaIleGly 446  
 Db 97 GTGATGGCGGTGCTGCTTCTTC-----CAGGTGACCTTCGACCTGACCTGCTGCC 150  
 Qy 447 AlaLeuMetThrValLeuGlyTyrSerLeuAsnThrLeuIlePheAspArgIle 466  
 Db 151 GCTGTGTCGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 210  
 Qy 467 ArgGluAsp-----ArgGlnAlaAsnLeuPheThrProMetHisValLeuVal 482  
 Db 211 CGCGAGAACTTCGCGTCTCGCAAGCCGATCTGGCGGAGAACCTGAGCATC----- 264  
 Qy 483 AsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThrLeuSer 502  
 Db 265 -----TCCACAGCCAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 318  
 Qy 503 ValLeuLeuMetLeuPheIleGlySerSerValPheAsnPheAlaPheIleMet 522  
 Db 319 GCTATCGCGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 378  
 Qy 523 ThrIleGlyLeuLeuGlyThrLeuSerSerLeuIleAlaProLeuLeuLeu 542  
 Db 379 TTCGTCGTGTAATGGCGGTACCTACTCTGTCGATCTACATCGCAACGTTGTTGATC 438  
 Qy 543 PheMet 544  
 Db 439 TGGCTG 444

RESULT 8  
 BH182894  
 LOCUS  
 DEFINITION 022\_B\_07-rev SmbAC1 Schistosoma mansoni genomic clone 022B07 5',  
 genomic survey sequence.  
 BH182894  
 VERSION BH182894.1 GI:16287528  
 KEYWORDS GSS.  
 SOURCE Schistosoma mansoni  
 ORGANISM Schistosoma mansoni  
 Eukaryote; Metazoa; Platyhelminthes; Trematoda; Digenea;  
 Strigoidida; Schistosomatidae; Schistosomatidae; Schistosoma.  
 1 (bases 1 to 602)  
 Le Paslier, M.-C., Pierce, R.J., Merlin, F., Hirai, H., Wu, W., Williams  
 D.L., Johnston, D., Loverde, P.T. and Le Paslier, D.  
 Construction and characterization of a Schistosoma mansoni  
 bacterial artificial chromosome library

# JOURNAL

Genomics 65 (2), 87-94 (2000)  
 20247247  
 MEDLINE  
 PUBMED  
 10783255

# COMMENT

Other\_GSSs: 022\_B\_07-21  
 Contact: Pierce RJ  
 INSERM U 167  
 Institut Pasteur de Lille  
 1 rue du Professeur A. Calmette, 59019-Lille, France  
 Tel: (33) (0)3 20877783  
 Fax: (33) (0)3 20877888  
 Email: Raymond.Pierce@pasteur-lille.fr  
 CNS sequencing ID=DG0AA022CA04BP1  
 Plate: 022 row: B column: 07  
 Seq primer: M13 reverse primer  
 Class: BAC ends  
 High quality sequence stop: 602.

# FEATURES

Location/Qualifiers  
 1. .602

/organism="Schistosoma mansoni"  
 /mol\_type="genomic DNA"  
 /strain="Puerto-Rican"  
 /db\_xref="taxon:6183"  
 /clone="022B07"  
 /sex="mixed"  
 /dev\_stage="cercariae"  
 /lab\_host="Biomphalaria glabrata"  
 /clone\_lib="SmbAC1"  
 /note="Vector: pBelobAC 11; Site 1: Hind III; Partially  
 Hind III digested and size-selected S. mansoni cercarial  
 DNA was ligated into Hind III digested pBelobAC 11 vector  
 and used to transform E. coli DH10B. The complete library  
 contains 23808 clones from 4 independent  
 singling-ligation-transformations. Average insert size  
 ranges from 70-127 kb and genome coverage is 7.9-Fold."  
 BASE COUNT 194 a 88 c 121 g 158 t 41 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 5,35e-12 Length: 602  
 Score: 203.50 Matches: 61  
 Percent Similarity: 46.63% Conservative: 29  
 Best Local Similarity: 31.61% Mismatches: 95  
 Query Match: 7.33% Indels: 8  
 DB: 28 Gaps: 3

US-09-868-987-14 (1-552) x BH182894 (1-602)

Qy 50 IleSerSerAspLeuGlyLysLysGlnCysThrGlnGlyIleIleSerAlaCysGly 69  
 Db 3 GTAGGACCTCTTTAGGACAAAAGCAATGATGCGAGGTATTTTCTTCATTAGTTGA 62  
 Qy 70 LeuAlaMetLeuValLeuMetSerValTyrArgPheGlyGlyValIleAlaSer 89  
 Db 63 TTCTTCTTWTATGTGTGGATGTTCTTATTACGTAGACGGGTGTTGTAATGCAAT 122  
 Qy 90 GlyAlaValLeuAsnLeuLeuIleTropAlaLeuLeuLeuTyrLeuAspAlaPro 109  
 Db 123 CGCGCTTTATTATAACGTTATTCTTATTGTTGTAATGCAAGTTTGTGTTTGA 182  
 Qy 110 LeuThrLeuSerGlyLeuAlaGlyIleValLeuAlaMetGlyMetAlaValAspAlaAsn 129  
 Db 183 TTAACATTGCCAGGTATTCAGGTATCGSATWAACATTAGGTACTGCGTGTGATGCAAC 242  
 Qy 130 ValLeuValPheGluArgIleArgGluPheLeuLeuSerGlnSerLeuLysSer 149  
 Db 243 ATCATTTAAAHGAAAGAGCAAGAAATWACGTGAAGGTAAAACTTTAGCAGAAGNA 302  
 Qy 150 ValGlu-----LysGlyTyrThrLysAlaPheGly-AlaIlePheAspSerLeu 167  
 Db 303 GTTGACATTTCAATGGTTGGAAAGAGCCANGCCGTAACAATAGTAGAAGCAACGTTAC 362  
 Qy 167 rThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGlyProIleLysGlyPheAl 187

Db 363 GCATATCTTAACGGGNATCANCCTTAANNNGNCNTNGAAACCGGACCCCAAGGANTTGC 422  
 Qy 187 aLeuThrLeuLeuGlyLeuPheSerMetPheThrAlaLeuPheMetThrLysPh 207  
 Db 423 GNNACAAANANCAAMAGGNANAAAAACNNCANNANANANANCAANCCCAAGCAAGAA 482  
 Qy 207 ePhePheMetLeuTrpMetAsnLys-----ThrGlnHisThrGlnLeuHisMe 223  
 Db 483 CNCMANCGACAAAGAGATGGCGAAGAGAGAACGGNAACCTTTTCAACTTCTATT---AC 539  
 Qy 223 tMetAsnLysPheValGlyLeuLysHisAspPheLeu 235  
 Db 540 GAAAAAAGTGTTCACAGGTTTCCATTTCGATTATC 576

RESULT 9  
 CNS0706R  
 LOCUS  
 DEFINITION T3 end of clone 022CA04 of library SMBAC1 from strain Puerto-Rican of Schistosoma mansoni, genomic survey sequence.

ACCESSION AL619845  
 VERSION AL619845.1 GI:16033987

KEYWORDS GSS.

SOURCE Schistosoma mansoni

ORGANISM Schistosoma mansoni

REFERENCE Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatidae; Schistosomatidae; Schistosoma.

AUTHORS 1 (bases 1 to 602)

Le Paslier, M.C., Pierce, R.J., Merlin, F., Hirai, H., Wu, W.,

Williams, B.L., Johnston, D., Loverde, P.T. and Le Paslier, D.

Construction and characterization of a Schistosoma mansoni

bacterial artificial chromosome library

Genomics 65 (2), 87-94 (2000)

JOURNAL 20247247

MEDLINE 10783255

REFERENCE 2 (bases 1 to 602)

Genoscope.

AUTHORS Direct Submission

TITLE Submitted (05-OCT-2001) Genoscope - Centre National de Recherche

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr

- Web : www.genoscope.cns.fr)

COMMENT Partially Hind III digested and size-selected S. mansoni cercarial

DNA was ligated into Hind III digested pBeloBAC 11 vector and used

to transform E. coli DH10B. The complete library contains 23808

clones from 4 independent sizing-ligation-transformations. Average

insert size ranges from 70-127 kb and genome coverage is 7.9-fold.

Location/Qualifiers

1..602

/organism="Schistosoma mansoni"

/mol\_type="genomic DNA"

/strain="Puerto-Rican"

/db\_xref="taxon:6183"

/clone="022CA04"

/clone\_lib="SMBAC1"

/note="end : T3"

BASE COUNT 194 a 88 c 121 g 158 t 41 others

ORIGIN

Alignment Scores:

Pred. No.: 5.35e-12 Length: 602

Score: 203.50 Matches: 61

Percent Similarity: 46.63% Conservative: 29

Best Local Similarity: 31.61% Mismatches: 95

Query Match: 7.33% Indels: 8

DB: 29 Gaps: 3

US-09-868-987-14 (1-552) x CNS0706R (1-602)

Qy 50 l1eSerSerAspLeuGlyLysGlnCysThrGlnGlyLeuLeuSerAlaCysGly 69

Db 3 GTAGGACCTTCTTAGGCAAAAGCAATTCATGCGAGTATTTATCTTATAGTTGGA 62

Qy 70 LeuAlaMetLeuLeuLeuMetSerValTyTrpArgPheGlyGlyValLeuAlaSer 89

Db 63 TTCTCTTTWTVATGTGTGTGGTGTCTTATTACGGTAGAGCGGTGTGTGTGTCACAAAT 122  
 Qy 90 GlyAlaValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 109  
 Db 123 CGCGCTTTATTATTAAACGTTATTATTCTTATTTGGTGTAAATGGCAAGTTTGGTTTGTGA 182  
 Qy 110 LeuThrLeuSerGlyLeuAlaGlyValLeuAlaMetGlyMetAlaValAspAlaAsn 129  
 Db 183 TTAACATTGCCAGTATTGAGGTATCGATWAACATTAGGTACTGTCAGTTGATGCCAAC 242  
 Qy 130 ValLeuValPheGluArgIleArgGluPheLeuLeuSerGlnSerLeuLysLysSer 149  
 Db 243 ATCATTAAGAAAGAGCAAGCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 302  
 Qy 150 ValGlu-----LysGlyTyThrLysAlaPheGly-AlaIlePheAspSerAsnLeuTh 167  
 Db 303 GTTGACATTTCAATGTTGGAAAGAGCGCCGCTTAACAAATAGTAGAAGCAACGTTAC 362  
 Qy 167 rThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGlyProIleLysGlyPheAl 187  
 Db 363 GCATATCTTAACGGGNATCANCCTTAANNNGNCNTNGGAACCGGACCCCAAGGANTTGC 422  
 Qy 187 aLeuThrLeuLeuGlyLeuPheSerSerMetPheThrAlaLeuPheMetThrLysPh 207  
 Db 423 GNNACAAANANCAAMAGGNANAAAAACNNCANNANANANANCAANCCCAAGCAAGAA 482  
 Qy 207 ePhePheMetLeuTrpMetAsnLys-----ThrGlnHisThrGlnLeuHisMe 223  
 Db 483 CNCMANCGACAAAGAGATGGCGAAGAGAGAACGGNAACCTTTTCAACTTCTATT---AC 539  
 Qy 223 tMetAsnLysPheValGlyLeuLysHisAspPheLeu 235  
 Db 540 GAAAAAAGTGTTCACAGTTCATTCGATTATC 576

RESULT 10

BZ560773/c

LOCUS

DEFINITION pacs2-164\_2779.x1 pacs2-164 Pseudomonas aeruginosa genomic clone

pacs2-164\_2779, genomic survey sequence.

ACCESSION BZ560773

VERSION BZ560773.1 GI:27179481

KEYWORDS GSS.

SOURCE Pseudomonas aeruginosa

ORGANISM Pseudomonas aeruginosa

REFERENCE 1 (bases 1 to 1124)

Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,

Burns, J.L., Kaul, R. and Olsen, M.V.

Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol., (2002) In press

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

FEATURES

Location/Qualifiers

1..1124

/organism="Pseudomonas aeruginosa"

/mol\_type="genomic DNA"

/strain="2-164"

/db\_xref="taxon:287"

/clone="pacs2-164\_2779"

/clone\_lib="pacs2-164"

/notes="clinical isolate 2-164 whole genomic shotgun library."

BASE COUNT 237 a 353 c 329 g 203 t 2 others

ORIGIN





QY 460 uillePheAspArgIleArgGluAsp-----ArgGlnAlaAenLeuPheT 476  
 Db 561 -CTGATCTTCAGCCGGTGGCAAGACTTTCGGCGTGTGGCGAAGCCGATCTGGTCG 503  
 QY 476 hrPrometHisValLeuValAenAspAlaLeuGlnLysThrPheSerArgThrValMetT 496  
 Db 502 AGAAGCTTGAACATC-----TCCACGAGCCAGACCCCTGTGGTACCATCGCCA 455  
 QY 496 hrThrAlaThrThrLeuSerValLeuMetLeuPheIleGlySerSerValP 516  
 Db 454 CTCCTGGTTCCACGCTGTGGCTATTCGGCGCGCTGTCTTCGCGCGCGCAATCTGT 395  
 QY 516 hAsnPheAlaPheIleMetThrIleGlyIleLeuGlyThrLeuSerSerLeuTyri 536  
 Db 394 TCGGTTTCTCCATCCCGCTGTTCGTCGGTGTATCGCGGTACCTACTCGTCGATCTACA 335  
 QY 536 leAlaProPheLeuLeuPheMet 544  
 Db 334 TCGCCAACGCTGATTTATGATCTGGCTG 309

RESULT 12  
 AZ303132 548 bp DNA linear GSS 06-MAR-2001

LOCUS GSSBr2079 Brucella abortus random genomic library Brucella  
 DEFINITION melitensis biovar Abortus genomic clone UU2079, genomic survey  
 sequence.

ACCESSION AZ303132.1 GI:10129343

VERSION GSS.

KEYWORDS

SOURCE

ORGANISM

Brucella melitensis biovar Abortus (Brucella abortus)

Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

Brucellaceae; Brucella.

REFERENCE 1 (bases 1 to 548)

AUTHORS Sanchez,D.O., Zandomeni,R.O., Cravero,S., Verdun,R.E., Pierrou,E.,

Faccio,P., Diaz,G., Lanzavecchia,S., Aguerro,F., Frasca,A.C.C.,

Andersson,S.G.E., Rosetti,O.L., Grau,O. and Ugalde,R.A.

Gene discovery through genomic sequencing of Brucella abortus

Infect. Immun. 69 (2), 865-868 (2001)

JOURNAL 21101034

MEDLINE 11159979

PUBMED

COMMENT

Contact: Siv Andersson

Small Genomes Sequencing Group

Department of Molecular Evolution, Uppsala University

Norbyvagen 18C, S-752 36, Uppsala, Sweden

Tel: 46-18-471-4379

Fax: 46-18-471-6404

Email: Siv.Andersson@ebc.uu.se

Sequences were basecalled with phred and vector was masked with

crossmatch (see <http://genome.washington.edu>). Sequences were then

trimmed from both ends to remove low quality bases and masked

VECTOR.

Class: shotgun.

Location/Qualifiers

1..548

/organism="Brucella melitensis biovar Abortus"

/mol\_type="genomic DNA"

/strain="2308"

/db\_xref="taxon:235"

/clone="UU2079"

/clone\_lib="Brucella abortus random genomic library"

/note="Vector: modified M13"

98 a 150 c 156 g 136 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 4,32e-08 Length: 548

Score: 168.00 Matches: 32

Percent Similarity: 66.29% Conservative: 27

Best Local Similarity: 35.96% Mismatches: 30

Query Match: 6.05% Indels: 0

DB: 28 Gaps: 0

US-09-868-987-14 (1-552) x AZ303132 (1-548)

QY 454 TyrSerLeuAenAsnThrLeuIlePheAspArgIleArgGluAspArgGlnAlaAen 473  
 Db 2 TATTCGCTGAACGATCGGTGTATATACCGGTGCGGAGAACTTCGCGGTAC 61  
 QY 474 LeuPheThrPrometHisValLeuValAenAspAlaLeuGlnLysThrPheSerArgThr 493  
 Db 62 AAGAGTCGCCATTCGCCGCGCATCATCGATCCCTCGAGGGCCAGACCTGTCGGAACG 121  
 QY 494 ValMetThrThrAlaThrThrLeuSerValLeuMetLeuPheIleGlyGlySer 513  
 Db 122 CTGTTGACCTCTTCGTCACCTTCCTGCGACATGTTCCGCTTTATGCTTTTGGCGGCTCG 181  
 QY 514 SerValPheAenPheAlaPheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSer 533  
 Db 182 GAAATTCGCATGTTTGGCGTCGCGCTCAGTCTCGGTATCATTTGTGGCGAGTTATTCTTCA 241  
 QY 534 LeuTyriLeAlaProProLeuLeuLeu 542  
 Db 242 ATCTTCATCGCAGCAGCCGCTGTTT 268

RESULT 13

CB852238

LOCUS

DEFINITION

UI-CF-FNO-aff-i-17-0-UI.s1 UI-CF-FNO Homo sapiens cDNA clone

UI-CF-FNO-aff-i-17-0-UI 3', mRNA sequence.

ACCESSION CB852238

VERSION CB852238.1

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 662)

AUTHORS

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

JOURNAL

MEDLINE

PUBMED

COMMENT

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

Contact: McCray, PB

McCray Lab

University of Iowa

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

Tel: 319 356 4866

Fax: 319 356 7171

Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics ([www.resgen.com](http://www.resgen.com)) or from Open Biosystems

([www.openbiosystems.com](http://www.openbiosystems.com)).

Seq primer: M13 FORWARD

PolyA=No.

Location/Qualifiers

1..662

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clones="UI-CF-FNO-aff-i-17-0-UI"

/tissue\_type="Human Lung Epithelial cells"

/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone\_lib="UI-CF-FNO"

/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a

modified polylinker; Site 1: EcoR I; Site 2: Not I;

UI-CF-FNO is a subtracted cDNA library derived from two

normalized Human lung epithelial cell libraries (EN1 and

DUI) The library was subtracted according to according to

Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. For additional information, contact:  
bento-soares@uiowa.edu  
TAG SEO=None found"

| BASE COUNT | 101 a | 215 c | 227 g | 117 t | 2 others |
|------------|-------|-------|-------|-------|----------|
| ORIGIN     |       |       |       |       |          |

Alignment Scores:

|                        |          |               |     |
|------------------------|----------|---------------|-----|
| Pred. No.:             | 6,478-07 | Length:       | 662 |
| Score:                 | 158.50   | Matches:      | 57  |
| Percent Similarity:    | 41.04%   | Conservative: | 46  |
| Best Local Similarity: | 22.71%   | Mismatches:   | 86  |
| Query Match:           | 5.71%    | Indels:       | 62  |
| DB:                    | 14       | Gaps:         | 9   |

US-09-868-987-14 (1-552) X CB852238 (1-662)

|     |    |   |     |
|-----|----|---|-----|
| 245 | Qy | SerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerVal  | 264 |
|     |    |   |     |
| 1   | Db | GGGGCGCGGTCTTCTCGTCAACCGTGGCGTCATTC                           | 42  |
|     |    |   |     |
| 265 | Qy | LeuGlyMetAspPheLysGlyTyrAlaPheThrPheAsnProLysGluHisGlyIle     | 284 |
|     |    |   |     |
| 43  | Db | ---TCATTCGAATTCAACGGCGGCACCGTCATC                             | 84  |
|     |    |   |     |
| 285 | Qy | SerAspValAlaGlnMetArgGlyLysValValHisLysGlnAlaGlyLeuSer        | 304 |
|     |    |   |     |
| 85  | Db | CCGAGGCGCGCGAGATCGAG---AAGACCCGCCACACCATCGAAGATGGGCTATGGC     | 141 |
|     |    |   |     |
| 305 | Qy | SerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLysIleTyrPhe     | 323 |
|     |    |   |     |
| 142 | Db | ---GAGGTCCAGGTGCAGAACTTCGGCTCTCTCGCGACGTCATGATCCGCTCGCG       | 195 |
|     |    |   |     |
| 324 | Qy | ---SerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerIleLeuLys        | 340 |
|     |    |   |     |
| 196 | Db | GTGCTGCGCGGAGAAAGACAGACCGAAGTGTGGCAAGGTCTTCGGTGAGCTCTCGCGC    | 255 |
|     |    |   |     |
| 341 | Qy | LeuThrIleMetSerTrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeu  | 360 |
|     |    |   |     |
| 256 | Db | GCCGAG-----GCCGGCGCGCTCAGCGAGCACCAAGNAGTCAGCGGC               | 297 |
|     |    |   |     |
| 361 | Qy | TyrGlyYasnSerLysArgAsnAlaLysPheTrpSerLysValSerSerLysLeuSerLys | 380 |
|     |    |   |     |
| 298 | Db | CAGGGC-----GAGCAGGTCAGCAAGCAGATCTCTCAAG                       | 330 |
|     |    |   |     |
| 381 | Qy | Lys-----  | 381 |
|     |    |   |     |
| 331 | Db | AAACCGAGGGCGGACGCCCATCGGTGTGCGGCTCCGAGGTGTGGTCCGGCGGCTC       | 390 |
|     |    |   |     |
| 382 | Qy | ---MetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleIleLeu        | 398 |
|     |    |   |     |
| 391 | Db | GGCGCGAGCTGGCCAGGACGCCCAAGGCCCTGGCCGTGCGGTGCCGGCATCATG        | 450 |
|     |    |   |     |
| 399 | Qy | LeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIle  | 418 |
|     |    |   |     |
| 451 | Db | ATCTACCTGGCCATCCGCTTCGAGTGGAAAGTTTCGCTGGCGGCATCATCGCAACCTG    | 510 |
|     |    |   |     |
| 419 | Qy | HisAsp-LeuLeuAlaThrCysAlaValLeuPheIleAlaHisPheLeuLysLysIle    | 438 |
|     |    |   |     |
| 511 | Db | CACGACNGTG-----GTCAATCATCCTGGGCTTCTTCGCTTCTTCAGTG             | 555 |
|     |    |   |     |
| 438 | Qy | eGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyr-SerLeuAsnAs | 458 |
|     |    |   |     |
| 556 | Db | GGAGTTCTCGTGGCGGTGCTGGCGCGGTCGCGGTCTCGGCTATTCGGTCAACGA        | 615 |
|     |    |   |     |
| 458 | Qy | nThrLeuIleIlePheAspArgIleArgGlu                               | 468 |
|     |    |   |     |
| 616 | Db | GTCCGTGGTGCATCTTCGACCGGGTGGCGGAG                              | 646 |

RESULT 14

|            |  |               |             |              |                 |
|------------|--|---------------|-------------|--------------|-----------------|
| CNS008NJ   | CNS008NJ   | 1101 bp       | DNA         | linear       | GSS 03-JUN-1999 |
| LOCUS      | Drosophila melanogaster  | genome survey | sequence T7 | end of BAC # |                 |
| DEFINITION | EACR17D15 of RPCI-98 library from Drosophila melanogaster (fruit |               |             |              |                 |

Db 916 ACCGCMATCTGCTCGCTGTTTCAGAGCGAGCTGACCTGCCBGGCATGCGGGCATTSTT 975  
 Qy 120 LeuAlaMetGlyMetAlaValAspAlaAsnValLeu 131  
 Db 976 CTGACCCCTGGCATKSGGTGATGCKACATCTCTTA 1011

## RESULT 15

BH372283/c

## LOCUS

BH372283 655 bp DNA linear GSS 10-DEC-2001  
 AG-ND-101E1.TR ND-TAM Anopheles gambiae genomic clone AG-ND-101E1,  
 genomic survey sequence.

## ACCESSION

BH372283

## VERSION

BH372283.1

## KEYWORDS

GI:17318408

## SOURCE

GSS.

## ORGANISM

Anopheles gambiae (African malaria mosquito)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;  
 Anopheles.

## REFERENCE

1 (bases 1 to 655)

## AUTHORS

Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren  
 C., Huff, E.R., Carlile, J.L., Black, K., Zhang, H.-B., Gardner, M.J.  
 and Collins, F.H.

## TITLE

Construction of a BAC library and generation of BAC end  
 sequence-tagged connectors for genome sequencing of the African  
 malaria mosquito *Anopheles gambiae*

## JOURNAL

Mol. Genet. Genomics 268 (6), 720-728 (2003)

## MEDLINE

22542063

## PUBMED

12655398

## COMMENT

Other GSSs: AG-ND-101E1.TF

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjloftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by  
 F.H. Collins and sequenced by The Institute for Genomic Research  
 (TIGR). The BAC library was generated from A. gambiae PEST strain  
 DNA. All DNA was extracted from newly hatched first instar larvae  
 to minimize the inclusion of DNA from microorganisms that inhabit  
 the gut. The DNA is derived from mixed sexes of larvae. The BAC  
 library was constructed at Texas A&M University BAC Center  
 University, College Station, Texas 77843-2123, USA using a HindIII  
 partial digest.

Seq primer: M13 Rev

Class: BAC ends.

## FEATURES

Location/Qualifiers

1..655

/organism="Anopheles gambiae"

/mol\_type="genomic DNA"

/strain="PEST"

/db\_xref="taxon:7165"

/clone="AG-ND-101E1"

/clone\_lib="ND-TAM"

/note="Vector: pECBAC1; Site\_1: HindIII"

197 a 129 c 107 g 222 t

## BASE COUNT

ORIGIN

## Alignment Scores:

Pred. No.: 4 98e-06 Length: 655

Score: 150.50 Matches: 49

Percent Similarity: 50.52% Conservative: 49

Best Local Similarity: 25.26% Mismatches: 77

Query Match: 5.42% Indels: 20

DB: 28 Gaps: 6

US-09-868-987-14 (1-552) x BH372283 (1-655)

Qy 138 GluGluPheLeuLeuSerGlnSerLysSerValGluLysGlyTyrThrYsala 157

Db 655 AAGAATTATTGTCAGGGAAGGAAATCCGTGAAGCATACACGATGTTTCAAGCAGCCT 596

Qy 158 PheGlyAlaIlePheAspSerAsnLeuThrThr--ValLeuAlaSerAlaLeuLeuPheP 177  
 Db 595 TAATCTCGGAATATTGACAGTACACACCAACACATACTAACAGCTGTTGTTATTATA 536  
 Qy 177 heLeuAspThrGlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSers 197  
 Db 535 TTTTCGGACAGGGACCAATCAAGGATTGCTGTAACCTTAATTATCGGTATTGATGA 476  
 Qy 197 erMetPheThrAlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrG 217  
 Db 475 CGTTCTTTTACTTCTGATTATTGTCGAGAGTAATGATCTTCAGCAGACTTCGAAAGGTA 416  
 Qy 217 InHisThrGlnLeu-----HisMetMetAsnLysPheValGlyIleLysHisAspP 234  
 Db 415 AAGAGATTTCAGTATGACATCTTCTCAAAAATCTTTTCAGAAATATCTGGATCGACT 356  
 Qy 234 heLeuArgGlyCysLysLysLeuTrpAla---ValSerGlySerValPheLeuLeuGlyC 253  
 Db 355 TTATCGGG-----AAAAGAAATGGTCTTACATCTTCTCTACTATATTGATGATTATCT 302  
 Qy 253 ysValAla-----LeuGlyPheGlyAlaTrpAsnSerValLeuGlyMetAspP 269  
 Db 301 GTATCGCATCTATTGTTACAAAAGGCTTC-----AAATTTGGGTTGACT 257  
 Qy 269 heLysGlyGlyTyrAlaPheThrPheAsnProLysGluHisGlyIleSerAspValAlaG 289  
 Db 256 TCAAAGGAGGAAGATTATGTTGTAAGATTGTAAGCCGGTTGTCGCTTCCGATATTC 197  
 Qy 289 InMetArgGlyLysValValHisLysLeuGlnGluAlaGlyLeuSerSerArgAspPheA 309  
 Db 196 AGGAAGAGTTGGCTCCGATATATAAACTAAT-GACGGTAAAAACGAAGCTGTAGAT--- 142  
 Qy 309 rgIleGlnThrPheGlySerSerGluLysIleLysIle 321  
 Db 141 --GTTAAACATTTGGTAACTCTAATCAGCTTAGAATA 106

Search completed: October 28, 2003, 18:15:52

Job time : 3282 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 28, 2003, 15:26:12 ; Search time 5913 Seconds  
(without alignments)  
3819.062 Million cell updates/sec

Title: US-09-868-987-14  
Perfect score: 2778  
Sequence: 1 MVSSPILNVLKHNASVSGK.....SLYIAPPLLLFMVRKNSK 552

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-Q=/Cgn2\_1/USPTO\_spool/US09868987/runat\_28102003\_152601\_23835/app\_query.fasta\_1.711  
-DB=GenEmbl -QFMT=fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09868987 @Cgn\_1\_1\_3635 @runat\_28102003\_152601\_23835 -NCPH=6 -ICPU=3  
-NO\_WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hcg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_dat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*

- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pin.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description                    |
|------------|--------|-------------|--------|-------|--------------------------------|
| 1          | 2730   | 98.3        | 12198  | 1     | AE002179 Chlamydia             |
| C 2        | 2730   | 98.3        | 14134  | 1     | AE001641 Chlamydia             |
| C 3        | 2730   | 98.3        | 110000 | 6     | AR310754_06 Continuation (7 of |
| C 4        | 2730   | 98.3        | 300550 | 1     | AP002547 Chlamydia             |
| 5          | 2088.5 | 75.2        | 304769 | 1     | AE016994 Chlamydia             |
| C 6        | 1937   | 69.7        | 9989   | 1     | AE001318 Chlamydia             |
| C 7        | 1924   | 69.3        | 12442  | 1     | AE002341 Chlamydia             |
| C 8        | 817.5  | 29.4        | 10046  | 1     | AE009114 Agrobacte             |
| C 9        | 817.5  | 29.4        | 10066  | 1     | AE008079 Agrobacte             |
| C 10       | 766.5  | 27.6        | 349619 | 1     | AP002996 Mesorhizo             |
| C 11       | 766    | 27.6        | 300600 | 1     | AP005952 Bradyrhiz             |
| C 12       | 762.5  | 27.4        | 14652  | 1     | AE004799 Pseudomon             |
| C 13       | 739.5  | 26.6        | 300861 | 1     | AE016777 Pseudomon             |
| C 14       | 736    | 26.5        | 7687   | 1     | AE012971 Chlorobiu             |
| 15         | 736    | 26.5        | 311249 | 1     | AE016860 Pseudomon             |
| 16         | 734.5  | 26.4        | 10963  | 1     | AE013703 Yersinia              |
| C 17       | 734.5  | 26.4        | 285050 | 1     | AJ414155 Yersinia              |
| C 18       | 725    | 26.1        | 282183 | 1     | CJ11168X4 AU139077 Campyloba   |
| C 19       | 722.5  | 26.0        | 3000   | 1     | U032710 Haemophilus            |
| C 20       | 719    | 25.9        | 13818  | 1     | U32710 Haemophilus             |
| C 21       | 719    | 25.9        | 110000 | 6     | AR274513_02 Continuation (3 of |
| C 22       | 713.5  | 25.7        | 11527  | 1     | AE010580 Fusobacte             |
| C 23       | 711    | 25.6        | 305961 | 1     | AE016937 Bacteroid             |
| C 24       | 709.5  | 25.5        | 15058  | 1     | AE000652 Helicobac             |
| C 25       | 709.5  | 25.5        | 19836  | 1     | AE001567 Helicobac             |
| C 26       | 707.5  | 25.5        | 14261  | 1     | AE004160 Vibrio ch             |
| C 27       | 699.5  | 25.2        | 300956 | 1     | AE016963 Coxsiella             |
| C 28       | 699.5  | 25.2        | 301442 | 1     | AE016798 Vibrio vu             |
| C 29       | 696    | 25.1        | 10384  | 1     | AE011297 Leptospir             |
| C 30       | 695    | 25.0        | 10611  | 1     | AE015748 Shawanell             |
| C 31       | 694    | 25.0        | 4435   | 1     | ECSECD F                       |
| C 32       | 694    | 25.0        | 10323  | 1     | AE005220 Escherich             |
| C 33       | 694    | 25.0        | 10577  | 1     | AE000147 Escherich             |
| C 34       | 694    | 25.0        | 13544  | 1     | AE015067 Shigella              |
| C 35       | 694    | 25.0        | 139818 | 1     | ECU82664 Escherichia           |
| C 36       | 694    | 25.0        | 292309 | 1     | AE016979 Shigella              |
| C 37       | 694    | 25.0        | 295741 | 1     | AP002551 Escherich             |
| C 38       | 693    | 24.9        | 300817 | 1     | AE016756 Escherich             |
| C 39       | 692.5  | 24.9        | 11540  | 1     | AE009509 Brucella              |
| C 40       | 691    | 24.9        | 20388  | 1     | AE008714 Salmonell             |
| C 41       | 691    | 24.9        | 268050 | 1     | AL627266 Salmonell             |
| C 42       | 691    | 24.9        | 300029 | 1     | AE016842 Salmonell             |
| C 43       | 689.5  | 24.8        | 11911  | 1     | AE014430 Brucella              |
| C 44       | 689    | 24.8        | 12201  | 1     | AE006057 Pasteurel             |
| C 45       | 687.5  | 24.7        | 301235 | 1     | AP005075 Vibrio pa             |

ALIGNMENTS

RESULT 1

AE002179 12198 bp DNA linear BCT 30-MAY-2000  
LOCUS Chlamydomophila pneumoniae AR39, section 15 of 94 of the complete  
DEFINITION genome.  
ACCESSION AE002179 AE002161  
VERSION AE002179.2 GI:8163380  
KEYWORDS  
SOURCE  
ORGANISM Chlamydomophila pneumoniae AR39  
Chlamydomophila pneumoniae AR39  
Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomophila.  
REFERENCE 1 (bases 1 to 12198)  
AUTHORS Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F.,  
White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T.,  
Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B.,  
Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J.,  
McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.  
Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
pneumoniae AR39  
Nucleic Acids Res. 28 (6), 1397-1406 (2000)  
JOURNAL 20150255  
MEDLINE 10584935  
PUBMED  
REFERENCE 2 (bases 1 to 12198)  
AUTHORS Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F.,  
White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T.,  
Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B.,  
Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J.,  
McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.  
Direct Submission  
TITLE Submitted (01-MAR-2000) The Institute for Genomic Research, 9712  
JOURNAL Medical Center Dr, Rockville, MD 20850, USA  
COMMENT On Jun 1, 2000 this sequence version replaced gi:7189108.  
FEATURES  
source  
1..12198  
/organism="Chlamydomophila pneumoniae AR39"  
/mol\_type="genomic DNA"  
/strain="AR39"  
/db\_xref="taxon:115711"  
/notes="synonym: Chlamydia pneumoniae AR39"  
145..1521  
/gene="CP0178"  
145..1521  
/gene="CP0178"  
/notes="similar to GB:L26051 SP:P33986 PID:415662;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="UDP-N-acetylglucosamine  
1-carboxyvinyltransferase"  
/protein\_id="AAF38052.1"  
/db\_xref="GI:7189109"  
/translation="MQIAQVGGCRLNGEVKVSAGKNAATKLLVASLLSDQKTLRNV  
PDGDVSLVELCKSLGAVSWDKTEVLEIYPTQCTRVPTFNSVNRPIILLIGA  
LLGRCPGVVPTVGGDAIGERTLNHFEGKQLGVQISSDSSGYAKAPRGKGNVI  
HLPPYSGATENILIAAHAKGRVTKVALEAIEDLVLFLOKAGADITDNDRTID  
IFGTGGVSDHTLIPKIEAAGFMAAVSGGRVFNRAKQELLIPFLKMLRSIGGG  
FLVSEGIREFORPLGVLETDHPGLTDQOPFVAVLLSOAGSSVHETVHEN  
RLGVHLGLQHWGECQFLPHOCLSTKACRYAIGNFPASAVIHGATPVAWSHLVDPDRA  
GFAYMAALAEAGGSIIENHLLDRGTYNNVGLKSLGAKIQIFDMEQBELTSPKS  
LAURDASL"  
complement(1475..3160)  
/gene="CP0179"  
complement(1475..3160)  
/gene="CP0179"  
/notes="similar to PID:1001350 PID:1001346; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="arginyl-tRNA synthetase"  
/protein\_id="AAF38053.1"  
/db\_xref="GI:7189110"  
/translation="MSTLLSILSVICSQAIAFPNLEDWAPETITSTKEHFGHYQCN  
DAMKLARVLKAPRAIEAIVAEIQPEPFLIEIAGAGFINFTFSPVFNQQLHEPKD  
ALKLGFQVSPKPIIIDFFSPNTAKDMHVGHLRSTIIGDSLARIFSVVGHVLRNLHI

GDWGTAFGLMITYLOENPCDSDLEDLTSLYKAYVCFTNDEBFKKRSQONVVALQAK  
DPOAIAWKEKICETSEKAKQKLYDILDIVVERGESFYNPFLPEIILEDKKGLLTVS  
NDAKVPHEAFSIPFNWQSGDGYVATTDLAMRYRIEEDHADKILIIIVTDLQCSLHF  
QLEATAIAAGLQPGIFSHVGLVDPOGKKLKTRSGENKVRKLELLODTAIEKAEA  
LREHPELTDEAIOERAPVIGNAIKYSDLSHRTSDYVFSPEKMLRFEGTAMFLLY  
AVVRIQIKRRLGISQLEGGPEIQEPAEELIALLTLRFPFALESTIKELCPHFLTD  
YLYNLTHKNGFFRDSHIQDSPYAKSRFLCALAEQVLAQVLAQVLAQVLAQVLAQV  
complement(3165..3803)  
/gene="CP0180"  
complement(3165..3803)  
/gene="CP0180"  
/notes="1-acyl-sn-glycerol-3-phosphate acyltransferase,  
putative; identified by match to PFAM protein family HMM  
PF01553"  
/codon\_start=1  
/transl\_table=11  
/product="1-acyl-sn-glycerol-3-phosphate acyltransferase,  
putative"  
/protein\_id="AAF73638.1"  
/db\_xref="GI:8163381"  
/translation="MIFRICKEFTVWAFSLFKLYGVYKKNFIKGPALIAVNHNSFL  
DPIALHMCVHECIYHLARSLFNIPMLKQWGFVVRQDEGSAAFKISRUPKRRKK  
LVYIPEGARSPDQQLQPGKVGIGMAAKSRVPIIPVYIRGTFAFNRHOKIPHWKTI  
TCVFTGPMYFDDIIQNPEIKNKETYQIITNQ"MNKIAELKAWYESGCKGDPV"  
complement(3800..4450)  
/gene="CP0181"  
complement(3800..4450)  
/gene="CP0181"  
/notes="similar to SP:P23863 PID:42839 GB:U00096  
PID:1651431 PID:1651438; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="cytidylate kinase"  
/protein\_id="AAF38054.1"  
/db\_xref="GI:7189111"  
/translation="MIITIDGPSGTGKSTTAKADHLHFNVCNTGMYRTIAYARLQ  
SPWATLPLTKFLEPSPFSTFATQPLESPFNGHLLTSELTTQEVANAASELSOLPEV  
RAFMQDLQRYAQLGNCVPEGRDMGSKVFPNADKLIFLTSSEVRAQRRLKDLPEGTL  
SPEQIAELVKRDAQADRAHPLVIPENGIVIDSSDLTIQVLEKILALLFRNEL"  
complement(4447..5373)  
/gene="CP0182"  
complement(4447..5373)  
/gene="CP0182"  
/notes="similar to GB:M1330 SP:P06466 PID:1208947  
PID:145476 GB:U00096; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="phosphatidate cytidyltransferase, putative"  
/protein\_id="AAF38055.1"  
/db\_xref="GI:7189112"  
/translation="MLNKNKFSKTKGAYGDLQFQVVVHSLVTLFLVLLYSDFPLTS  
FALGITATCGAVGTVEYSSMAKAMHVPSTFSAIGSFLFALSPLSIRWGHSLPGF  
PDALWTLIIWVWVSIIPRVKRTIGALQSGVTLFSLYGVIGIRLPLVLYSPHTL  
QEPYLIGWASLIATTKGADIFGVFGKAGFNKKAIOISVKNKTVGVGACIGLHT  
ISFIPLQIPTRFASYPFPFPAIIFELGALGKIFGFGDIIESIFKRAHAKKSNKUKA  
VCGMLTDLSLLSLPIAYLFLITQSKFEIG"  
complement(5374..6159)  
/gene="CP0183"  
complement(5374..6159)  
/gene="CP0183"  
/notes="similar to PID:1208946 GB:U00096 SP:Q47675  
PID:1552751 PID:1786371; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="undecaprenyl pyrophosphate synthetase"  
/protein\_id="AAF38056.1"  
/db\_xref="GI:7189113"  
/translation="MTYALINDPVDLSLATNNAESKFFSLQRLPNHVAIIMDGNRRWY  
KHRECECHTHTSGHYGAKVILNLAVIDLGIKVLITYFTSTENFGPKKEIOEIF  
NIFYTQDLQKPLYMENIEICIGDLSKLPKGIQTKINHVSMTASPSRLELVLAVN

YGGDELVAFKKLHVILNKKISDDLSLSSLSYSLTSLGTLDPDLLIRTGEMRVS  
NFLWQIAYTYLTITLWPDFTQDLFEAINVYQRRRGCK"  
complement (6214. .6363)  
/gene="CP0184"  
complement (6214. .6363)  
/gene="CP0184"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAF38057.1"  
/db\_xref="GI:7189114"  
/translation="MPKKPLYSFEVRLTKNLFLETPICLNRIQPLPTGMAGEKSFVI  
FVLGK"  
6637. .7737  
/gene="CP0185"  
6637. .7737  
/gene="CP0185"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAF38058.1"  
/db\_xref="GI:7189115"  
/translation="MKFPRIISDLIPTQVIMWRGGNVHVNAQNLPKKILGGVL  
ACFGLALGCAAFAGVCTIFPCIGLMILGLVLLGPAYLQYSGKSRFEPLEPREK  
VFEPDINWGLCSLLQSWKKIRPGCYHPGQVCEIPEGSOEIVTKIOPKSDRSTSI  
FLIQMDQIALRQSGKSLSRKTFADPSVSSLLSEIQEEQQYLDPPKVISWSSD  
QASDRTHPKSAIYNIISDAQEPQGRYIDYAKFTVLDQIGDNPVKKHTIYVLT  
PILGVPDLKPEENLKLQAPLQSAEPLYSAEQVAKMRBEKQDSIRIKFIETDPTPTS  
LVFSPHSSTPHSVTPISLSGFVGEQSYTFA"  
7917. 12125  
/gene="CP0186"  
7917. 12125  
/gene="CP0186"  
/note="similar to GP:3220156; identified by sequence  
similarity; putative"  
/codon start=1  
/transl\_table=1  
/product="secDF protein, putative"  
/protein\_id="AAF38059.1"  
/db\_xref="GI:7189116"  
/translation="MKQVKNFALICVFALALYVLPCLYVAKPLDKKIDGNEAE  
HLIKFTQAQVRKDVIPRVSAILSLHSGHQHQAIPDIIVSVRFKGEDADEAE  
GNLVGERNPKSLARLHVGVYSREHDDHVIQVASSINTSVESDFSEVYSSENEQE  
MASSILQVRYSACTFPKQKDCSCYSPIWETAPKEQLLQYAKNLSGFEVFSRLSAF  
COQSFSSNQDLAFSLRSLNSNDAAIDVEDQKLLKSVYETLSQTACTIRSDCPYIEG  
LRLDCSSLPFSSIEYCPKPKIFLTLHSDLLAORTSLSKQRLDPSRLAVKOKL  
SKNLTVOVEDYNNGFSFOWMDKTOGKIILQGERLQGI AEHLTALTILHRPAESCDL  
IPENPVFCROPRESEARGCYIFSPNTDCKHFSKGSVILLKGLRSIVAKYQOQGGKE  
LOSFEKQNLNYSHTSEALSWTIGEDQVLEIRHPLQOFLDVMWGEVPIQEGECAPL  
EYKQIDQRLATVNOIEKNRQSLVRWHEQYHAKMSDMLQRLSAPTYQNLFLFNKL  
LNMRFSGRGNILRLIGIDFVGQRQLLLSFQKHQKQLTKEDILKVSDELCAKLNKLG

gene

CDS

gene

CDS

gene

CDS

## Alignment Scores:

Pred. No.: 6, 09e-239 Length: 12198  
Score: 2730.00 Matches: 551  
Percent Similarity: 99.28% Conservative: 0  
Best Local Similarity: 99.28% Mismatches: 1  
Query Match: 98.27% Indels: 3  
DB: 1 Gaps: 0

US-09-868-987-14 (1-552) x AE000179 (1-12198)

Qy 1 MetValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20  
Db 10464 ATGGTCAGAGCCCTATTITTAACCTCCCATTTGAAATATCATGCAGGTCTCAGGGAAA 10523  
Qy 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40  
Db 10524 TTATCCACCGTGAAGTCAGCAAACTCGCCTCAGATTTTAAATCTGGAGCGATGCTTTT 10583

Qy 41 ValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysGlnCysThr 60  
Db 10584 GTTCCCGAGGTTCTCAGTGAAGACCGATCTCTCTCATCTTGGGAAAAAACAATGTACA 10643  
Qy 61 GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValThr 80  
Db 10644 CAAGGCATTATCTCAGCATGCTGTGGCTTGCAATGCTTATTGTTTGTATGAGCGTATAT 10703  
Qy 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrp 100  
Db 10704 TATAGATTGGAGCGCTCATCGCTTCGGAGCTCTCTTCTGAATCTCTTCTTATCTG 10763  
Qy 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120  
Db 10764 GCAGCTCTACAGTATTGGATGCGCCACTCACCTTGTGAGGACTCGCTGGGATGTTCTT 10823  
Qy 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPhe 140  
Db 10824 GCTATGGGATGGCGGTAGATGCAATGTTCTTGTATTGCAAGAAATCCGAGAGGAATTT 10883  
Qy 141 LeuLeuSerGlnSerLeuLysLysSerValCylulysGlyTyrThrLysAlaPheGlyAla 160  
Db 10884 TTAFTGTCTCAAGTCTTAAAAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCC 10943  
Qy 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180  
Db 10944 ATTTTTCATCTAACTTGTACTACAGTATTGGCCTCAGCACTTCTTTCTTCTAGATACA 11003  
Qy 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200  
Db 11004 GGGCTATTAAAGGTTTGCTTTGACATTGATTTAGGAATTTTCTCTCAATGTTTACG 11063  
Qy 201 AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisThrGln 220  
Db 11064 GCTCTTTTCATGACTAAATTTTCTTCTCATGCTGTGGATGAATAAGATCCCAACATACACAG 11123  
Qy 221 LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys 240  
Db 11124 TTGCATATGATGAATAAGTTCTGGGGGATTAAGCATGATTTCTTGAGAGGATGCANAAA 11183  
Qy 241 LeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla 260  
Db 11184 CTTTGGGCTGTTCTGGAAGTGTCTTTCTTTAGTGTGCGTGTCTCTCGGCTTTGGAGCC 11243  
Qy 261 TrpAsnSerValLeuGlyMetAspPheLysGlyClyTyrAlaPheThrPheAsnProLys 280  
Db 11244 TGGAAATTCGTTTTGGGAATGGATTTTAAAGGAGGGGTATGCCCTTTTAAATCCAAA 11303  
Qy 281 GluHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGlu 300  
Db 11304 GAGCATGGCATCAGCGATGTTGCTCAATGGTGGCAAGTTGTGTCATAACTACAGGAA 11363  
Qy 301 AlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGlyLysIleLys 320  
Db 11364 GCTGGTCTTTCTCTAGAGACTTCGGTATTCAACATTTTGGATCTTCAGAAAAAGATCAA 11423  
Qy 321 IleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLys 340  
Db 11424 ATCTATTATTAGTGATAAGCTTTTAAAGCTATATAAGCTATATAAGCAGATACGAGCCCTCTCTCTATAA 11483  
Qy 340 sLeuThrIleMetSerTrpArg-TyrCysGlyIleValValValArgAsnArgProArgPheL 360  
Db 11484 ATTAACGATCATGAGCTGGCGTTAGCTGTGGGATTTGTTGTGAGAAACAGGCTAGATTTTC 11543  
Qy 360 eutYrGlyAsnSerLysArgAsnAlaLys-PheTrpSerLysValSerSerLysLeuSer 379  
Db 11544 TCTACGGAACCTCTAAACGAAACGCAAAATTTTGGTCAAGGTAAAGCAAGCAAACTATCG 11603  
Qy 380 LysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleIleLeuLeu 399  
Db 11604 AAGAAATGCGTTATACGGCGACCATCGGCTTTTAGGAGCTTTTGGCAATCATCTTGCTC 11663

```
Qy 400 TyrValSerLeuAArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuLeuHis 419
Db 11664 TATGTAGTTTGGCTTTGCAATGCAATATGCTTTAGTGGCGTATGCGCTTTAATTCAT 11723
Qy 420 AspleuLeuAlaThrCysAlaValLeuPheIleAlaHisPhePheLeuLysLysIleGln 439
Db 11724 GACCTTTTGGCTACTGTCGTCAGTCTTGTATTATAGCACATTTCTTTTGAAGAAAATTCAA 11783
Qy 440 IleAspleuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnThr 459
Db 11784 ATAGATTTCGAAGCAATTTGGTGCTTTATGACTGTATTGGGTATTCATTAAACAATACT 11843
Qy 460 LeuIleLeuPheAspArgIleArgGluAspArgGlnAlaAsnLeuPheThrProMetHis 479
Db 11844 TTGATCATTTTGTATGCTGTTTCGTAAGATGCCAAGCAGACCTGTTTACCCCTATGCAT 11903
Qy 480 ValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThr 499
Db 11904 GTTTTAGTTAATGATGCCCTTCAAAAGACGTTACGCCGACGGTAATGACAAACAGCTACA 11963
Qy 500 ThrLeuSerValLeuLeuMetLeuLeuPheIleGlySerSerValPheAsnPheAla 519
Db 11964 ACTCTATCAGTTTGTAAATGCTTTTGTATTATAGCGCGCTCCTCTGCTTTAATTTTGA 12023
Qy 520 PheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProPro 539
Db 12024 TTTTATTAGCATAGGATCTTCTAGGAACCTTTATTCGTTCTTTATATATGCACCACT 12083
Qy 540 LeuLeuLeuPheMetValArgLysGluAsnArgSerLys 552
Db 12084 CTGTTGTTGTTTATGTCGTTACGTAAGAAAATCGCTCAAAA 12122

RESULT 2
AE001641/c
LOCUS AE001641 14134 bp DNA linear BCT 01-DEC-2000
DEFINITION Chlamydia pneumoniae section 57 of 103 of the complete genome.
ACCESSION AE001641 AE001363
VERSION AE001641.1 GI:4376855
KEYWORDS Chlamydia pneumoniae pneumoniae CWL029
SOURCE Chlamydia pneumoniae pneumoniae CWL029
ORGANISM Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE 1 (bases 1 to 14134)
AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.W., Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE 99206606
PUBMED 10192388
REFERENCE 2 (bases 1 to 14134)
AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
FEATURES
source
1. .14134
/organism="Chlamydia pneumoniae pneumoniae CWL029"
/mol_type="genomic DNA"
/strain="CWL029"
/db_xref="taxon:115713"
/complement(158. .1924)
/gene="recJ"
/notes="synonym: CPn0563"
/complement(158. .1924)
/gene="recJ"
/codon_start=1
/transl_table=11
/product="ssDNA Exonuclease"
/protein_id="AAD18703.1"
/db_xref="GI:4376856"
/translation="MTNSDNASAGLLWAHPKEDPAFLGMIKEFHPPTVAQIFSR
GFQTIQIHKFLYSLSSLYDPGLFDMSKAVERLLARDKEHVMYIGDSVDGWTG

gene
/notes="synonym: CPn0566"
8009. .8761
/gene="yaeS"
/codon_start=1
/transl_table=11
/product="YaeS family hypothetical protein"
/protein_id="AAD18706.1"
/db_xref="GI:4376859"
/translation="NSLATNNAESKFPSPQLRPNHVAIIMDGNRRMYKXHEECGHTH
TSCHYVGAFLNPLNLAVDLGIKVLTYLTFSTENFGPKBEIQIENFYQLDKQL
PYLMENEICLRICIGDLSLPGIKGIQTKINHVSMTASFSLRLVLAVNMGKDELVRAF

gene
/notes="synonym: CPn0565"
8009. .8761
/gene="CPn0565"
/codon_start=1
/transl_table=11
/product="CT449 hypothetical protein"
/protein_id="AAD18705.1"
/db_xref="GI:4376858"
/translation="MKFPRISLDLIPTQMVIWRGGNVHYVVPNAQNPKKILGGVL
ACFGALLGCAAFAGVCTIPCCIGMLGLVLLGFAVLOYSKMSRERPLFRFTK
VPEKINMLGCLSLLOSNNKIPGCGYHPGCPQVEICEGSEIVTKIPKKSDRNTSI
FLIQEMDQALRGQIEKSLRSLRTAIDPSVSSLSERQEEQQYLDPKVLSWSSSED
QASDRTHPKSAIYNISDAAPQPCRYATIGLGAIAIILYLSURFEMQVAFS
PILGVPDALPKEENKLLSQAFLYSAEQVAKRMREKQDSIRIKFTFTDPTSPTS
LYFSPHSTPHSVTPISLSGFGVQESYTA"
```

```
gene
/notes="synonym: CPn0564"
complement(2010. .6218)
/notes="secD/secF"
/genes="secD/secF"
/codon_start=1
/transl_table=11
/product="Protein Export Proteins SecD/SecF (fusion)"
/protein_id="AAD18704.1"
/db_xref="GI:4376857"
/translation="MKQKVRNFALIIICVFALALVYVPTCLYAKPDKKIDGNEAE
HIKFTKQQRKQDVIPRVSAILSSHLFGHIQHPALPDIVSVFRGGEADDFI
GNLVGEPNVPKSKARLHVGYSRBDDHVIQVASSINTSLVESDFSVYSSENOE
MASSILQRYISACTTTPKQDCSCYPSIWETAPKEQLQYAKNLSGDFVFSRLSAF
COQSSNODRLAFLSLSDAAIDVEIQKLLKSVYETLSQACIRSLDCPYTEG
LRIDCESLFFSSEIYCPKPKRIPLTLHSULLAORTSLSKRQRLDFDRLAWEKOKL
SKNLTVOVEDYNGFSFOMDKDTQGGKIILOGERLLOGIAEHLTALTILHRAAESCDL
IENPVFCRPRESEAFGCIYFSPNTDCKHFSKGSVYILLKGLSIVAKYQGGCKE
LOSFKDLQNLNCFSTHSAISWTIGEDQVLEIRHPLQQFLDVGWGGFVYKGGCAFL
EYKDIQDLATVNIQEKRSQDLVRWHEQYHAKSMQDLERLSAPIYONIFLENMK
LNMRFSGRENILRLIGIDFVGQRQLLSFKHQKQLTDKEDILKVSDBLCARLNKLG
VSEIELRREGDYIHLSPGSSSTISSLEILGTSKMGSHVYVNERFSSYASRYEQRFLD
YLMFTSOAQGKTSPEIINTFASALFNEEDVDPSPVHEAITKLKSEGLASPSGCEPS
TOLDTFMSIALGKDAEQKANPLVIVFRNALDGLASLKDIREFAAGEGYVLFNSVD
TSFKKAAEKLSPSTFHTWTSAYCOGISGTANGQYSANRGMMAVVDGYWVSSDIL
NPLKNHASVSGKFTHREYVSLASDLKSGAMSFVPEVSEETISSDLGKKQCTQGIIS
ACCGLAMLVMSVYVFGGVTASGAVLLNILLIWAALOYLDAPLTLSGLAGTILAMG
MAVDANLVFERIRREFLLSQSLKSKSVKGYTKAFCAI FDSNLTTLASALLPFLDTG
PIKFALTLILGIESMFTALEMTKFFEMLMNKTQHTOLHNMNKFVGIKHDFLRGCK
KLWAVSGSVLLGCVAGLFGGAWNSVDMDFGKGYYAFTFPKEHGIDVAQMRGKVHVK
LQEAGLSRDFRIQTFGSEKIKIYFSDKALSYTKADTSLSPKINDHEALVGLLUSE
TGIDFTSETNETONFWSKSKSKMRYQATIGLGAIAIILYLSURFEMQVAFS
AVCALIHDLATCAVLFAHFKKIQIDLQALGALMTVLGYSLANTLIIIFDRIDR
QANLTPHVLVNDALQKTFSTVTMTATTLSVLLMLLFI GGSSVFNFAPIMTIGILL
GTLSLIYAPLLLEFWRKENSK"
complement(6398. .7498)
/genes="CPn0565"
/notes="CPn0565"
/genes="CPn0565"
/codon_start=1
/transl_table=11
/product="CT449 hypothetical protein"
/protein_id="AAD18705.1"
/db_xref="GI:4376858"
/translation="MKFPRISLDLIPTQMVIWRGGNVHYVVPNAQNPKKILGGVL
ACFGALLGCAAFAGVCTIPCCIGMLGLVLLGFAVLOYSKMSRERPLFRFTK
VPEKINMLGCLSLLOSNNKIPGCGYHPGCPQVEICEGSEIVTKIPKKSDRNTSI
FLIQEMDQALRGQIEKSLRSLRTAIDPSVSSLSERQEEQQYLDPKVLSWSSSED
QASDRTHPKSAIYNISDAAPQPCRYATIGLGAIAIILYLSURFEMQVAFS
PILGVPDALPKEENKLLSQAFLYSAEQVAKRMREKQDSIRIKFTFTDPTSPTS
LYFSPHSTPHSVTPISLSGFGVQESYTA"
```





```
QY 241 LeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuLeuGlyPheGlyAla 260
Db CTTTGGGCTGTTCTGGAAGTGTCTTTCTTTAGTTGGTGTCTCTCGGGTTGGAGCC 2892
QY 261 TrpAsnSerValLeuGlyMetAspPheLysGlyGlyTyrAlaPheThrPheAsnProLys 280
Db TGGAAATTCGGTTTGGGAATGGATTTAAAGAGGGGTATGCTTACCTTTAAATCCAAAA 2832
QY 281 GluHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGlu 300
Db GAGCATGGCATCAGGATGTTGCTCAAAATGGTGGCAAAAGTTGTGCATAAACTACAGAA 2772
QY 301 AlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLys 320
Db GCTGGTCTTTCTCTAGAGACTCCGTATTCAACAATTTGGATCTTCAGAAAAAGATCAAA 2712
QY 321 IleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLys 340
Db ATCTATTTTAGTGATAAGCTTTAAGCTATATACTAAAGCAGATACGAGCCTCTCTCTAAA 2652
QY 340 sLeuThrIleMetSerTrpArg-TyrCysGlyIleValValArgAsnArgProArgPheL 360
Db ATTAAACGATCATAGCTGGCGTTAGCTGGGATTTGTTGTCAAGAACAGGCTTAGATTTTC 2592
QY 360 euTyrGlyAsnSerLysArgAsnAlaLys-PheTrpSerLysValSerSerLysLeuSer 379
Db TCTACGGAAACTCTAAACGAAACGCAAAATTTTGGTCAAAAGGTAAGCAGCAAACTATCG 2532
QY 380 LysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeu 399
Db AAGAAATATGCGTTATCAGGCGACCATCGGGCTTTTAGGAGCTTTGGCAATCATCTTGCTC 2472
QY 400 TyrValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHis 419
Db TATGTGAGTTTGGCTTTGAATGCAATATGCTTTCACTGCGGTATGCGCTTTTAATTCAT 2412
QY 420 AspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPheLeuLysLysIleGln 439
Db GACCTTTTGGCTACCTGTGCAGTCTTGTTTATAGCACATTTCTTTTGAAGAAATTCAA 2352
QY 440 IleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThr 459
Db ATAGATTTGCAAGCCATTTGGTCTTTAATGACTGTATTTGGGGTATTCATTAAACAATACT 2292
QY 460 LeuIleIlePheAspArgIleArgGluAspArgGlnAlaAsnLeuPheThrProMetHis 479
Db TTGATCATTTTGTATCGTATTCGTGAAGATCGCCAGCGAACCTGTTTACCCCTATGCAT 2232
QY 480 ValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThr 499
Db GTTTTAGTTAATGATGCCCTTCAAAAGAGGTTTCAGCCGACGGTAATGACAAACAGCTACA 2172
QY 500 ThrLeuSerValLeuLeuMetLeuLeuPheIleGlySerSerValPheAsnPheAla 519
Db ACTCATTCAGTTTGTATTAATGCTTTTGTATAGCGGGCTCTCTGTCTTTAATTTTGCA 2112
QY 520 PheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProPro 539
Db TTTATTATGACCATAGGATTTCTTAGAAACTTTATCGTCTCTTTATATTTGACACACCT 2052
QY 540 LeuLeuLeuPheMetValArgLysGluAsnArgSerLys 552
Db CTGTGTGTTTATGTTATGGTCCGTAAGAAAAATCGCTCAAAA 2013
RESULT 3
AR310754_06/c
WPCOMMENT
Sequence split into 13 fragments LOCUS AR310754 Accession AR310754
Fragment Name Begin End
AR310754_00 1 110000
AR310754_01 100001 210000
AR310754_02 200001 310000
AR310754_03 300001 410000
```

```
AR310754_04 400001 510000
AR310754_05 500001 610000
AR310754_06 600001 710000
AR310754_07 700001 810000
AR310754_08 800001 910000
AR310754_09 900001 1010000
AR310754_10 1000001 1110000
AR310754_11 1100001 1210000
AR310754_12 1200001 1230025
Continuation (7 of 13) of AR310754 from base 600001 (AR310754 Sequence 1 from patent US
Alignment Scores:
Pred. No.: 1,11e-237 Length: 110000
Score: 2730.00 Matches: 551
Percent Similarity: 99.28% Conservative: 0
Best Local Similarity: 99.28% Mismatches: 1
Query Match: 98.27% Indels: 3
DB: Gaps: 0
US-09-868-987-14 (1-552) x AR310754_06 (1-110000)
QY 1 MetValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20
Db 61906 ATGGTCAGCAGCCCTATTTTAAACGTCCTCCATTCGAAAAATCATGCCAGTGTCTCAGGGAAA 61847
QY 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40
Db 61846 TTTACCCACCGTAGTGAGCAAACTCGCCCTCAGATTTTAAATCTCGAGCGATGTCTTT 61787
QY 41 ValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysLysGlnCysThr 60
Db 61786 GTTCCCGAGGTTCTCAGTGAAGAGACGATCTCTTCTGATCTTGGGAAAAACAATGTACA 61727
QY 61 GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80
Db 61726 CAAGGACATATCTCAGCATGCTGTGCTTGGCAATGCTTATTTTGTATGAGCGTATAT 61667
QY 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuLeuLeuLeuTip 100
Db 61666 TATAGATTGGAGCGTCATCGCTTCGGAGCGTGTCTCTGAAATCTTTTGTCTTATCTGG 61607
QY 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120
Db 61606 GCAGCTCTACAGTATTTGGATGGCCACTCACCTTGTGAGGACTCGCTGGGATTTGTCTT 61547
QY 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPhe 140
Db 61546 GCTATGGGATGGCCGTAGATGCAAAATGTTTGTATTCGAAAGAAATCCGAGAGGAATTT 61487
QY 141 LeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160
Db 61486 TTAATGTTCAAGTCTTTAAAAAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCC 61427
QY 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180
Db 61426 ATTTTGTATTAACCTTGACTACAGTATTTGGCCTCAGCACTCTTTTCTTCTTCTTAGATACA 61367
QY 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200
Db 61366 GGGCCTTATTAAGGGTGTCTTTGACATTTGATTTTATAGGAATTTTCTTCTCAATGTTTACG 61307
QY 201 AlaLeuPheMetThrLysPhePheMetLeuTipMetAsnLysThrGlnHisThrGln 220
Db 61306 GCTCTTTTTCATGACTAAAAATTTTCTTCATGCTGTGGATGATGATTAAGACCCCAACATACACAG 61247
QY 221 LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys 240
Db 61246 TTGCATATGATGAATAAGTTTCGTGGGATAAAGCATGATTTCTTGAGAGGATCAAAAAA 61187
QY 241 LeuTrpAlaValSerCysSerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla 260
Db 61186 CTTTGGGCTGTCTTCTGGAAGTGTCTTTTCTTTTAGGTTTCGTTGCTCTCGGGTTTGGAGCC 61127
```

QY 261 TrpAsnSerValLeuGlyMetAspPheLysGlyGlyTyrAlaPheThrPheAsnProLys 280  
Db 61126 TCGAATTCGTTTGGGAATGATTTAAAGAGGAGGTATGCTTTACCTTTAAATCCAAA 61067  
QY 281 GluHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGlu 300  
Db 61066 GAGCATGCGATCAGCGATGTGCTCAAAATGGTGGCAAAAGTTGTGCATAAACTACAGGAA 61007  
QY 301 AlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLys 320  
Db 61006 CTGCTGCTTCTTCTAGAGACTTCGTTATTAACATTTGGATCTTCAGAAAAGATCAAA 60947  
QY 321 IleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLys 340  
Db 60946 ATCTATTATTAGTATAAAGCTTTAAGCTATATACTAAAGCAGATACGAGCTCTCTCTATAA 60887  
QY 340 IleuThrIleMetSerTyrArg-TyrCysGlyIleValValArgAsnArgProArgPheL 360  
Db 60886 ATTACGATCATGAGCTGGCTGTAGCTGTGGATTTGTTCAGAAACAGGCTTAGATTTC 60827  
QY 360 euTyrGlyAsnSerLysArgAsnAlaLys-PheTyrSerLysValSerSerLysLeuSer 379  
Db 60826 TCTACGGAACCTTAACAGCAACGCAAAATTTTGTCAAGGTAAGCAGCAAACTATCG 60767  
QY 380 LysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeu 399  
Db 60766 AAGAAATGCGTTATCAGCGCACCATCGGCTTTTAGGAGCTTTGGCAATCATCTTGCTC 60707  
QY 400 TyrValSerLeuArgPheGluTyrGlnTyrAlaPheSerAlaValCysAlaLeuIleHis 419  
Db 60706 TATGTGAGTTTGGCTTTTGAATGGCAATATGCTTTTCAGTGGCGTATGCGCTTTAATTCTAT 60647  
QY 420 AspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPhePheLeuLysLysIleGln 439  
Db 60646 GACCTTTTGGTACCTGTGCGAGCTTGTATTAGCACATTTCTTTTTCAGAAATTCCTAA 60587  
QY 440 IleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnThr 459  
Db 60586 ATAGATTTGCAAGCCATTGGTCTTAAATGACTGTATTGGGTATTCATTAACAATACT 60527  
QY 460 LeuIleIlePheAspArgIleArgGluAspArgGlnAlaAsnLeuPheThrProMetHis 479  
Db 60526 TTGATCATTTTGTATGCTGATTCGTGAAGATCGCCAAGCGAACCTGTTTACCCTTATGCTAT 60467  
QY 480 ValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThr 499  
Db 60466 GTTTTAGTTAATGATGCCCTTCAAAAGAGCTTCAGCCGACGGTAAATGACACACTACA 60407  
QY 500 ThrLeuSerValLeuLeuMetLeuLeuPheIleGlyGlySerSerValPheAsnPheAla 519  
Db 60406 ACTCTATCAGTTTCTGTTAATGCTTTTGTATTATAGCGCGCTCCTCTGCTTTAATTTTGA 60347  
QY 520 PheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProPro 539  
Db 60346 TTTATTATGACCATAGGATTCCTCTAGGAACCTTTATGCTCTCTTTATATATGACCACT 60287  
QY 540 LeuLeuLeuPheMetValArgLysGluAsnArgSerLys 552  
Db 60286 CTGTTGTTTATGCTCGTAAGAAATTCCTCAAAA 60248  
RESULT 4  
LOCUS AP002547/c  
DEFINITION Chlamydomophila pneumoniae J138 genomic DNA, complete sequence,  
section 3/4.  
ACCESSION AP002547 AB033786 AB033787 AB033816 AB033817 BA000008  
VERSION AP002547.2 GI:10176693  
KEYWORDS  
SOURCE Chlamydomophila pneumoniae J138  
ORGANISM Chlamydomophila pneumoniae J138  
Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomophila.  
REFERENCE 1  
AUTHORS Shirai,M., Hirakawa,H., Ouchi,K., Tabuchi,M., Kishi,F., Kimoto,M.,  
Takeuchi,A., Nishida,J., Shibata,K., Fujinaga,R., Yoneda,H.,  
Matsushina,H., Tanaka,C., Furukawa,S., Miura,K., Nakazawa,A.,  
Ishii,K., Shiba,T., Hattori,M., Kuhara,S. and Nakazawa,T.  
Comparison of outer membrane protein genes omp and pmp in the whole  
genome sequences of Chlamydia pneumoniae isolates from Japan and  
the United States  
J. Infect. Dis. 181 Suppl 3, S524-S527 (2000)  
20298986  
10839753  
REFERENCE 2  
AUTHORS Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,  
Shiba,T., Ishii,K., Hattori,M., Kuhara,S. and Nakazawa,T.  
Comparison of whole genome sequences of Chlamydia pneumoniae J138  
from Japan and CWL029 from USA  
Nucleic Acids Res. 28 (12), 2311-2314 (2000)  
20330349  
10871362  
REFERENCE 3 (bases 1 to 300550)  
AUTHORS Shirai,M.  
TITLE Direct Submission  
JOURNAL Submitted (04-JUL-2000) Mutsunori Shirai, Yamaguchi University  
School of Medicine, Department of Microbiology; 1-1-1  
Minamikogushi, Ube, Yamaguchi 755-8505, Japan  
(E-mail:mshirai@po.cc.yamaguchi-u.ac.jp, tel:81-836-22-2227,  
Fax:81-836-22-2415)  
COMMENT On or before Sep 15, 2000 this sequence version replaced  
gi:6172298, gi:6172300, gi:6172396, gi:6172398, gi:8978889.  
FEATURES  
source  
1. 300550  
/organism="Chlamydomophila pneumoniae J138"  
/mol\_type="genomic DNA"  
/strain="J138"  
/db\_xref="taxon:138677"  
/note="synonym:Chlamydia pneumoniae (strain J138)"  
/gene="dapF"  
complement(21..794)  
/gene="dapF"  
complement(21..794)  
/codon\_start=1  
/transl\_table=11  
/product="diaminopimelate epimerase"  
/protein\_id="BAA98725.1"  
/db\_xref="GI:8978890"  
/translation="MAPYSPSTISKYFIYSCAGNRFLLGETLPEVEDRVFLCOBTRVD  
GFLVLPSSCADAQLIIFNSDGRPTMCGNGLRCAIAHLASOKGSDISVSTDGLYS  
GYEYSDRVLDVMTLADRASVHLESRPDLPEKVCVHTGVPHAVILPEISTDL  
SILGPFURYQTFTSPGVNVNFVQLGHGQJRVYTYERGVEGTAACTGALASALVV  
SNSYGMKESIQTHTWGGELMTVSQNGRVYLGQSVTRDL"  
/gene="clpP\_1"  
complement(763..1338)  
/gene="clpP\_1"  
complement(763..1338)  
/codon\_start=1  
/transl\_table=11  
/product="CLP protease"  
/protein\_id="BAA98726.1"  
/db\_xref="GI:8978891"  
/translation="MADGEVHKLRDIIIEKELLEARRVFFSEPVTEKASDAIKLWL  
ELKDPGRPIVFINSPGSGVADGAVWDQIKMLTSPVITVTGLAASMGSVLSUCAAP  
GRRFATPHSRIMIHQPSIGGPIGTQATDLDIHAREILKTKARIIDVYVEATNQPRDI  
EKAIIDRMWMTANEAKDFGLDGLILFSNDL"  
complement(1358..2917)  
/gene="glyA"  
complement(1358..2917)  
/gene="glyA"  
complement(1358..2917)  
/codon\_start=1  
/transl\_table=11  
/product="serine hydroxymethyltransferase"  
/protein\_id="BAA98727.1"  
/db\_xref="GI:8978892"  
/translation="MLKVFEKFKFAIVEITKVAVVSLLLHKPLENAGSKKGOSLAS  
TAYLAALDHLNAPPSIGERIIDELKQSRHLKMIASENYSLSQVAMGNLLPTDKYC  
EGSPFKRFYSCCNVDAIEWCVETAKELFAADCAACVQPHSGADANLLAVMAILTHKV

QCPAVSKLGYKTVNELTEEEYTLKAEWSSCVCLGSLNSGGHLTHGVRNLNVMSKLM  
WCDVNPDPTECFDYAEISLAKKEYKPKVLIAGYSYSSRYNLAASDLVILPVAELDPFNLTPTNST  
RVDMAHFAVGFVDEENPIPIYADIVITTHKTLRGPRGLVLAIRKQAEBCGSVL  
CPLMVEGPHVIAAKTVALKEALSDFKKYAHQVWNNARRLAERFISHGLRITLGT  
DNHMWIDGLSGISGKIAEDILSSVGIANNRSLPSDAIGKMDTSIGRLGTALATL  
GMGIDEMEVEADIIIVKLRNLRLSCHVEGSSKKNKGLPEALQAQEAADRVRNLLRPP  
LYPIDLEALV  
3107. .3778  
/gene="CP10522"  
3107. .3778  
/gene="CP10522"  
/codon\_start=1  
/transl\_table=11  
/product="CT433 hypothetical protein"  
/protein\_id="BAA98728.1"  
/db\_xref="GI:8978893"  
/translation="MTLYLGINOKTARKYQAHVLPILTPPYAKSTPONKRALQFLPQ  
ATHVILSPBSSTHLFLSRMSSLKATLTKTYLCIGESTKERLLSLFGQVYVATQ  
EIARGIFPLQALPSSARILYPHSSLARPVIREFLNRFYFSPHYVYPRKLKQNI  
LSYKKIIFTSPTVRAPAKIFPRFPKTYWCQGRMTLQBFQKPSQKQVSLLETGK  
SRTPSP"  
3843. .4175  
/gene="CP10523"  
3843. .4175  
/gene="CP10523"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAA98729.1"  
/db\_xref="GI:8978894"  
/translation="MASSATPGPDGTAPSLFPPTATRPYRNFKALFVITIAIALVWIAL  
IATTAIGLCIHPLCSFIFUTAIPLFYISHYARNVYIALDVVPDHSKLQWRS  
HSPFSDR"  
4223. .5302  
/gene="CP10524"  
4223. .5302  
/gene="CP10524"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAA98730.1"  
/db\_xref="GI:8978895"  
/translation="MSGPSTESSQSVLSVPRDKIAPKKQFTIAKISTAILASL  
ALGAVAGISLTVLGVFLALDITALLSFVTVFLVYHQMTKVSNNQKVLQNFK  
PLGMAQKNDVCYQNMYPNNHNPFRKVAIQTDASQFPQFTFLGLRVIKKNST  
GIIPNVPPTNLINTNLSTILYSLTKDKSVMDCKQREGPAKEDPSPTEVVR  
VKLPNEALDOTNLNLSSAEKSLTPTFLGHVCGPKSEELPNQOEYVROALLAVENCL  
KALESAAVAALPLFTSVYVEPPEILPKEGTFYWDNQIQAFCCKALLDAIQNTAIR  
YQRLSLVILQDFNTIESQSRSEE"  
5642. .6406  
/gene="CP10525"  
5642. .6406  
/gene="CP10525"  
/codon\_start=1  
/transl\_table=11  
/product="CT398 hypothetical protein"  
/protein\_id="BAA98731.1"  
/db\_xref="GI:8978896"  
/translation="MHDALLSILAIQELDIKMLRMVKEKHEKELAKVOSLKDIIR  
KVQKELEMENLKTQIRDGNRQIEISEQINKLENQQAAYKKMDEFNALTQEMTANK  
ERSLEHQLSDLMQKQGEDLVSLKESLASTENSSSVIEKEIFESIKKINEGKAL  
LEQRTLEKHTNPELLSIYERLLNNKDRVYVPIENRVCSGCHVILTPQHENLVKRD  
RLIFCEHCSRILYQESQVNAQENSTAKRRRRRAAV"  
complement(6827. .7816)  
/gene="kpsF"  
complement(6827. .7816)  
/gene="kpsR"  
/codon\_start=1  
/transl\_table=11  
/product="GutQ/kpsF family sugar-P isomerase"  
/protein\_id="BAA98732.1"  
/db\_xref="GI:8978897"  
/translation="MPSFMISTDVCQDILKQKEAVDFFQAFQPKENMQLAEBKILGH

SGWFFSGVGVKSCCVARKLVATLQSLSERALFSPVDLLHGDGLGLVSPGDIVCLFSSK  
TCQMIIFDVLAMLLFHSRGVSLTSTYGNHPSPQGVKMGKVKDFPKPTEVPFCHLG  
DKVSLGVLEFSAVCGCVCTVDPQRLMGIFTDGLRRSLASYGVEVLSLSLRKVMTA  
NPRCITEDSDIAIALQLMESSPVAVLPLVDLNEENRHVTGLLHMHHTLAKAGLL"  
complement(7843. .9024)  
/gene="sucB\_2"  
complement(7843. .9024)  
/gene="sucB\_2"  
/codon\_start=1  
/transl\_table=11  
/product="dihydrolipoamide succinyltransferase"  
/protein\_id="BAA98733.1"  
/db\_xref="GI:8978898"  
/translation="MIFEPFPKIGETSSGGSIVRWLKNLGDHVARDEPLIEVSTDKI  
ATLPSPKAGRLVPCNVEGDEVASGDVLGLIELEISEADDESTCPLTSCCTKSEA  
GSSSSWFSPPASVLSAOREGIGLNDLQIKTAKTGKGRVTRDPLEAVISQOVSIPE  
IFQGVNRPMSPLRAIASLSKSSDEVPASLVVDVDTDLNLIISGERQRLDTH  
GVKLITTSIVQCLAQTLKQFPLNGSLDGTTLVMKKSNNVGVANLREGVVIVRYEV  
NCODRGLVSIKALADLSSRLNKLDPSVDQSVTVNFGMTGALIGMPIIRYPEV  
AIIIGICTIQRVVVRDDDSLAKRMVYVTLTDFHRLVDLGIYGSFELTSLKRLSEVTM  
G"  
complement(9038. .10282)  
/gene="gitT"  
complement(9038. .10282)  
/gene="gitT"  
/codon\_start=1  
/transl\_table=11  
/product="glutamate symport"  
/protein\_id="BAA98734.1"  
/db\_xref="GI:8978899"  
/translation="MKLWKKIFIGLVFVGLSLVEDKAIFFPKPIDGIIFLNLISMVYV  
PLVFCSMVLGIASIDMKKLRIGIKSVGLYGTTLAIVIGLCFAMIFSPGNCDFPA  
QAQSMDSAVTVIDSNTAAVFLSIIAQVFPSPNVRSPFAEGNIQIIFIIFLGLIALRL  
SCGERGPVERFIDGSEIMLRVMNMTSPAP/GVGASMAWISGNHGLVWLQKPTII  
AYYLACLPHATLVFGGLVRFCKMSFKSLSNMDAISCAVSTASSATLPTVMRCVS  
KNLGVSAEVSFVLPLGATVNMNGTAIFQGMMAVFAQAYNCPLSUSSLLLVVATF  
SASGEGELSPYSIKQESVETT"  
complement(10282. .11379)  
/gene="ycaH"  
complement(10282. .11379)  
/gene="ycaH"

Alignment Scores:  
Pred. No.: 4,19e-237 Length: 300550  
Score: 2730.00 Matches: 551  
Percent Similarity: 99.28% Conservative: 0  
Best Local Similarity: 99.28% Mismatches: 1  
Query Match: 98.27% Indels: 3  
DB: 1 Gaps: 0

US-09-868-987-14 (1-552) x AP002547 (1-300550)

Qy 1 MetValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20  
Db 50924 ATGGTCAGCAGCCCTATTTAAACGTCCTCATGAAATAATCATGCCAGTGTCTCAGGAAA 50865

Qy 21 PreThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40  
Db 50864 TTTACCCACCGTAAGTACGACAAATCGCCTCAGATTTTAAATCTCGAGCGATGCTCTTT 50805

Qy 41 ValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysLysGlnCysThr 60  
Db 50804 GTTCCCGAGGTTCTCAGTGAAGAGACGATCTCTTCTGATCTTGGGAAAAACAATGTACA 50745

Qy 61 GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80  
Db 50744 CAAGGCATTATCTCAGCATGCTGTGGCTTGGCAATGCTTATTGTTTGTATGAGCGTATAT 50685

Qy 81 TyrArgPheGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuLeuTyr 100  
Db 50684 TATAGATTGGAGCGCTCATCGCTTCGGAGCTGTTCTTCTGAATCTTTTGTCTATCTGG 50625

Qy 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120  
 Db 50624 CGAGCTCTACAGTATTGGATGGCGCACTACCTTGTGAGNCTCGCTGGGATGTTCTT 50565  
 Qy 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPhe 140  
 Db 50564 GCTATGGGAGTGGCGGTAGATGCAATGTTCTTGTATTGCAAGAATCCGAGAGGAATTT 50505  
 Qy 141 LeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160  
 Db 50504 TTAITGTCCTCAAGAGTCCTTAATAAATCTGTAGAAAAAGGATATACCAAGGCTTTTGGAGCC 50445  
 Qy 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180  
 Db 50444 ATTTTGTATTCTAACTTGCATACAGTATTGGCTCAGCACCTCTTTTCTTCTAGATACA 50385  
 Qy 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200  
 Db 50384 GGGCCTATTAAAGGGTTTGGCTTTGACATTGATTTTGGAAATTTTCTCTTCAATGTTTACG 50325  
 Qy 201 AlaLeuPheMetThrLysPhePhePheMetLeuTyrMetAsnLysThrGlnHisThrGln 220  
 Db 50324 GCTCTTTTCATGACTAAATTTTCTTCATGCTGTGGATGAATAAGCCCAACATACACAG 50265  
 Qy 221 LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys 240  
 Db 50264 TTGCATATGATGAATAAGTTCGTGGGATAAAGCATGATTTCTTGAGAGGATGCATAA 50205  
 Qy 241 LeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla 260  
 Db 50204 CTTTGGGCTGTTTCTGGAAGTGTCTTTCTTTAGGTGCGTGTCTCGGGTTTGGAGCC 50145  
 Qy 261 TrpAsnSerValLeuGlyMetAspPheLysGlyTyrAlaPheThrPheAsnProLys 280  
 Db 50144 TGGAAATCCGTTTGGGAATGATTTTAAAGAGGGTATGCTTTTACCTTTTAATCCAAA 50085  
 Qy 281 GluHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGlu 300  
 Db 50084 GAGCATGGCATCAGCGATGTTGCTCAATGCGTGGCAAGTTGTGCTAAACTACAGAA 50025  
 Qy 301 AlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLys 320  
 Db 50024 GCTGCTCTTCTCTAGAGACTTCGCTATCAACATTTGGATCTTCAGAAAAGATCAA 49965  
 Qy 321 IleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLys 340  
 Db 49964 ATCTATTTTAGTGATAAAGCTTTAAGCTATATACTAAAGCAGATACGAGCTCTCTCTAAA 49905  
 Qy 340 sLeuThrIleMetSerTrpArg-TyrCysGlyIleValValArgAsnArgProArgPheL 360  
 Db 49904 ATTAACGATCATGACTGCGGTAGCTGTGGATTTGTCAGAAAACAGGCTAGATTTC 49845  
 Qy 360 euTyrGlyAsnSerLysArgAsnAlaLys-PheTrpSerLysValSerSerLysLeuSer 379  
 Db 49844 TCTACGGAACCTCTAAACGAAACGCAAAATTTTGGTCAAAAGGTAAAGCAGCAAACTATCG 49785  
 Qy 380 LysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeu 399  
 Db 49784 AAGAAATGCGTTATCAGCGCCACCATCGGGCTTTTAGGAGCTTTTGGCAATCATCTTGCTC 49725  
 Qy 400 TyrValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHis 419  
 Db 49724 TATGTAGTTTGGCTTTTGAATGGCAATATGCTTTTCAGTGGCGTATGCGCTTTAATTCAT 49665  
 Qy 420 AspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPheLeuLysLysIleGln 439  
 Db 49664 GACCTTTTGGCTACCTGTGCACTCTGTTTATAGCACATTTCTTTTTCAGAAATTCAA 49605  
 Qy 440 IleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuLeuGlyTyrSerLeuAsnAsnThr 459  
 Db 49604 ATAGATTTGCAAGCAATTTGGTCTTTAATGACTGTATTGGGTATTTCATTAACAATACT 49545

Qy 460 LeuIleIlePheAspArgIleArgGluAspArgGlnAlaAsnLeuPheThrProMetHis 479  
 Db 49544 TTGATCATTTTGTATCGTATTCGTGAAGATCGCAAGCAACCTGTTTACCCCTATGCAT 49485  
 Qy 480 ValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThr 499  
 Db 49484 GTTTTAGTTAATGATGCCCTTCAAAGACGTTTACGCCACGCGTAATGACAACAGCTACA 49425  
 Qy 500 ThrLeuSerValLeuMetLeuLeuPheIleGlyLysSerSerValPheAsnPheAla 519  
 Db 49424 ACTCTACAGTTTGTATAGCTTTTGTATAGCGGCTCTCTGTCTTTAATTTTGTCA 49365  
 Qy 520 PheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProPro 539  
 Db 49364 TTTATTATGACCATAGGATTTCTTAGGAACCTTTATCGTCTCTTTATATGACCACT 49305  
 Qy 540 LeuLeuLeuPheMetValArgLysGluAsnArgSerLys 552  
 Db 49304 CTGTTGTTGTTTATGTCGTAAGAAATCGCTCAAAA 49266  
 RESULT 5  
 AEO16994 304769 bp DNA linear BCT 14-APR-2003  
 LOCUS Chlamydomophila caviae GPIC section 1 of 4 of the complete genome.  
 DEFINITION AEO16994 AEO15925  
 ACCESSION AEO16994.1 GI:29834116  
 VERSION  
 KEYWORDS  
 SOURCE Chlamydomophila caviae GPIC  
 ORGANISM Chlamydomophila caviae GPIC  
 Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomophila.  
 REFERENCE 1 (bases 1 to 304769)  
 AUTHORS Read T.D., Myers G.S., Brunham R.C., Nelson W.C., Paulsen I.T., Heidelberg J., Holtzapple E., Khouri H., Federova N.B., Carty H.A., Umayam L.A., Haft D.H., Peterson J., Beanan M., White O., Salzberg S.L., Hsiao R.C., McClarty G., Rank R.G., Bavoil P.M. and Fraser C.M.  
 TITLE Genome sequence of Chlamydomophila caviae (Chlamydia psittaci GPIC): examining the role of niche-specific genes in the evolution of the Chlamydiaceae  
 JOURNAL Nucleic Acids Res. 31 (8), 2134-2147 (2003)  
 MEDLINE 22569155  
 PUBMED 12682364  
 REFERENCE 2 (bases 1 to 304769)  
 AUTHORS Read T., Myers G., Brunham R., Nelson W., Paulsen I., Heidelberg J., Holtzapple E., Khouri H., Federova N., Carty H., Umayam L., Haft D., Peterson J., Beanan M., White O., Salzberg S., Hsiao R.C., McClarty G., Rank R., Bavoil P., and Fraser, C.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-OCT-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
 FEATURES  
 source  
 1. 304769  
 /organism="Chlamydomophila caviae GPIC"  
 /mol\_type="genomic DNA"  
 /isolate="GPIC"  
 /db\_xref="taxon:227941"  
 57..1052  
 /gene="hemb"  
 /locus\_tag="CCA00001"  
 57..1052  
 /gene="hemb"  
 /locus\_tag="CCA00001"  
 /notes="similar to GB:L16784, SP:P33395, and PID:290384; identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="delta-aminolevulinic acid dehydratase"  
 /protein\_id="AAP04754.1"  
 /db\_xref="GI:29834117"  
 /translation="MSSLALLRRPRNRKTAIRDLVAETALLPQDFICPFFVKEGKN ICEIESLPGVYRWSIDLKLEIERLCSLKAVILFPVPSDLKDAYGVSSNPKN I LCKSIYEVKKAPDLQCVISDIALDPYTHGHGDMDOCEVINDSESVIRFGNIATLHAE MGADIVAPSDMDGRVHIRAKLDQSGWTNTLILSYSVKVVASSLYNFRDLGSLHQS



/product="exodeoxyribonuclease V, gamma subunit"  
 /protein\_id="AAP04761.1"  
 /db\_xref="GI:29834124"

## Alignment Scores:

|                        |           |               |        |
|------------------------|-----------|---------------|--------|
| Pred. No.:             | 2,03e-178 | Length:       | 304769 |
| Score:                 | 2088.50   | Matches:      | 421    |
| Percent Similarity:    | 83.81%    | Conservative: | 50     |
| Best Local Similarity: | 74.91%    | Mismatches:   | 72     |
| Query Match:           | 75.18%    | Indels:       | 19     |
| DB:                    | 1         | Gaps:         | 3      |

US-09-868-987-14 (1-552) x AE0016994 (1-304769)

|    |        |  |        |
|----|--------|--|--------|
| Qy | 1      | MetValSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys    | 20     |
| Db | 189026 | GTGGTCAGCGATCTGTAATTAAGTCTCTTTAAAGATACGCTAGTGTTCTTGGAAAA     | 189085 |
| Qy | 21     | PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe | 40     |
| Db | 189086 | TTTTACATCTGTAAGTAACCGCTAGCTACGGATTATAAATCAGGATCTATGTCCTTT    | 189145 |
| Qy | 41     | ValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysGlnCysThr    | 60     |
| Db | 189146 | GTTCTGAGGTTTTAAAGTGAAGATTTATCTCTCTGAATTAGGCAACAAACGATCA      | 189205 |
| Qy | 61     | GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr | 80     |
| Db | 189206 | CAAGGATATCTCTATATGTTTAGTCTGCTGTTTGATCACTTTTGTAGCGGTTTTAT     | 189265 |
| Qy | 81     | TyrArgPheGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTyr       | 100    |
| Db | 189266 | TACAAATTCGGAGGAGTTATTGCTCGGAGCTGTTATCTTAAATCTTCTATTGATTGG    | 189325 |
| Qy | 101    | AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu | 120    |
| Db | 189326 | GCAGCCCTACAGTATCTTGATGCCCCCTAACTCTCAGCGGATTAGCGGGAATTGATTG   | 189385 |
| Qy | 121    | AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluLysPhe | 140    |
| Db | 189386 | CCATGGGAATGGCTGTAGATGCTTACGTTGTTTTCGAAAGAAATTCGGGAAGAGTAT    | 189445 |
| Qy | 141    | LeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGlyAla | 160    |
| Db | 189446 | TTATTATCGGAAGCTTACCAGCTCTGTAAGACGAGGTATAAAAAGCTTTCAGCGCT     | 189505 |
| Qy | 161    | IlePheAspSerAsnLeuThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr    | 180    |
| Db | 189506 | ATCTTTGATTGCAATTTGACCAACAATATTAGCTTCGCTACTTTTGGTTTATAGATACG  | 189565 |
| Qy | 181    | GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr | 200    |
| Db | 189566 | GGACCAATTAAGGAATTCGCTGACTTTAATCCTAGGAGTTTCTTCTCAATGTTTACT    | 189625 |
| Qy | 201    | AlaLeuPheMetThrLysPhePheMetLeuTyrMetAsnLysThrGlnHisThrGln    | 220    |
| Db | 189626 | GCATTATTATGACGAGGTTTCTTCTCATGTGATGATTAACCAACCAAGAACTCAG      | 189685 |
| Qy | 221    | LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys | 240    |
| Db | 189686 | CTGCATATGATGAATAAGTTCTTCGGCATCAAGCATGATTTCTTAAAGAGTGCAAAAGA  | 189745 |
| Qy | 241    | LeuTyrAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla | 260    |
| Db | 189746 | CTTTGGATGGTATCCGGGAGGCTTATTGTTTGGGATCGGTGCGTTAGGTTTCGGGGCA   | 189805 |
| Qy | 261    | TyrAsnSerValLeuGlyMetAspPheLysGlyTyrAlaPheThrPheAsnProLys    | 280    |
| Db | 189806 | TGGAAATTCGTTTGGGAATGGAATTTTAAAGTGGTTACGCTTTCACATTTAGACATGCT  | 189865 |
| Qy | 281    | GluHisGlyIleSerAspValAlaGlnMetArgGlyLysValHisLysLeuGlnGlu    | 300    |
| Db | 189866 | GATCAGAAAGCTGTAGATGTTGTACAGTTCGTATAAAGTTAGGCGGATAAATTTAAGCAG | 189925 |

|    |        |  |        |
|----|--------|--|--------|
| Qy | 301    | AlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLys   | 320    |
| Db | 189926 | GTAGGCTTATCTCTAGAGACTTCAAAATTAAGACGTTAGACTCTTCTGAGAGATAAAA     | 189985 |
| Qy | 321    | IleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSer            | 337    |
| Db | 189986 | ATATACTTCACTCAAAACGCAATTAAC-----CGAGTACAAGCTTCTGAGAAACAG       | 190036 |
| Qy | 338    | -----LeuLeuLysLeuThrIleMetSerTyrPheGlyCysGly                   | 350    |
| Db | 190037 | GTATTCGAAACCCCTGATCCTAATTTATCTTATGATGGCATACTTCTTCATACAGGA      | 190096 |
| Qy | 351    | IleValValArgAsnArgProArgPheLeuTyrGlyAsnSerLysArgAsnAlaLysPhe   | 370    |
| Db | 190097 | ATAGACACTTCTTCAGAG-----AGCTTTAAAGATACCAGCAATTTTC               | 190138 |
| Qy | 371    | TrpSerLysValSerSerLysLeuSerLysLysMetArgTyrGlnAlaThrIleGlyLeu   | 390    |
| Db | 190139 | TGGCTTAAAGTTAGCGACAGTTCTCTAATAAGATCGGTCAACAAGCTTGTATAGCGTTG    | 190198 |
| Qy | 391    | LeuGlyAlaLeuAlaIleIleLeuLeuTyrValSerLeuArgPheGluTyrGlnTyrAla   | 410    |
| Db | 190199 | CGCGGAGCCTTAGCGATCTTCTTATGTAGCTTGCCTTGTGAATGCGCTTATGCA         | 190258 |
| Qy | 411    | PheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeuPheIle   | 430    |
| Db | 190259 | TTTACTGCTGTTTCCGCAATGATTCATGATCTTGTCTGCTACTCGCGCGTATTAGTAGCC   | 190318 |
| Qy | 431    | AlaHisPhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThr   | 450    |
| Db | 190319 | ACACATTTCTTTTGCAGAAAGATACAAATGATTACAAAGCAGTCGGAGCTTGTATGACG    | 190378 |
| Qy | 451    | ValLeuGlyTyrSerLeuAsnAsnThrLeuIleIlePheAspArgIleArgGluAspArg   | 470    |
| Db | 190379 | GTTTTAGGGTATTCGTTTAAATAATACCTTAATCATCTTTCGATCGTATTCGAGAAGATCGT | 190438 |
| Qy | 471    | GlnAlaAsnLeuPheThrProMetHisValLeuValAsnAspAlaLeuGlnLysThrPhe   | 490    |
| Db | 190439 | CAGGAAAAATTTATTTACCTATGCCGATTTAATAAAGCAGCAGCTGCAAAAAACCTTA     | 190498 |
| Qy | 491    | SerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMetLeuPheIle      | 510    |
| Db | 190499 | GGAAGAACAGTAAATGACCAACAGCAGCAACACTGTCGGTATTATTAATTTTACTGTTGT   | 190558 |
| Qy | 511    | GlyClySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeuGlyThr   | 530    |
| Db | 190559 | GGAGCGGCTTCAATCTTTAACTTTGCTTTCATCATGCAATAGGTATTCTTTTGGGCACA    | 190618 |
| Qy | 531    | LeuSerSerLeuTyrIleAlaProLeuLeuLeuPheMetValArgLysGluAsnArg      | 550    |
| Db | 190619 | CTATCATCTCTATACATAGCACCCCTCTCTTTTATTCATGTTCTGTAAGAGGAAAGA      | 190678 |
| Qy | 551    | SerLys 552   |        |
| Db | 190679 | ACAGG 190684   |        |

RESULT 6  
 AE001318/c  
 LOCUS  
 DEFINITION Chlamydia trachomatis section 45 of 87 of the complete genome.  
 ACCESSION AE001318 AE001273  
 VERSION AE001318.1 GI:3328875  
 KEYWORDS  
 SOURCE Chlamydia trachomatis  
 ORGANISM Chlamydia trachomatis  
 Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
 REFERENCE  
 1 (bases 1 to 989)  
 AUTHORS Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R., Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q., Koonin,E.V. and Davis,R.W.  
 TITLE Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis



JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
2 (bases 1 to 9989)  
Kelman, S., Mitchell, W., Marathe, R., Lammel, C.J., Fan, J., Hymen, R.W.,  
Olinger, L., Grinnell, J., Davis, R.W. and Stephens, R.S.  
Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
Nat. Genet. 21 (4), 385-389 (1999)  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
3 (bases 1 to 9989)  
Stephens, R.S., Korman, S., Lammel, C.J., Fan, J., Marathe, R.,  
Aravind, L., Mitchell, W.P., Olinger, L., Tatusov, R.L., Zhao, Q.,  
Koonin, E.V. and Davis, R.W.  
Direct Submission  
Submitted (20-MAY-1998) Program in Infectious Diseases, University  
of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA  
FEATURES  
source  
1. 9989  
/organism="Chlamydia trachomatis"  
/mol\_type="genomic DNA"  
/strain="D/UW-3/CX"  
/db\_xref="taxon:813"  
complement (105. .371)  
/gene="omcA"  
/note="synonym: CT444"  
complement (105. .371)  
/gene="omcA"  
/codon\_start=1  
/transl\_table=11  
/product="9kDa-Cysteine-Rich Lipoprotein"  
/protein\_id="AAC68043.1"  
/db\_xref="GI:3328876"  
/translation="MKKTALLALCSVVSSCCRIDVCCFEDPCAPIQCSPCESKKK  
DVGGCNCNGYVPACKPCGGDTHQDAKHPQARGIPVDGKCRQ"  
708. .932  
/gene="CT444.1"  
708. .932  
/gene="CT444.1"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAC68048.1"  
/db\_xref="GI:3522897"  
/translation="MKRFFLIVATALLTQGCSEYPPKREYILAKCSPFLFKHAQTL  
DFSLAQAVFEMTFTRAPPCVSSLELSON"  
complement (929. .2449)  
/gene="gltX"  
/note="synonym: CT445"  
complement (929. .2449)  
/gene="gltX"  
/codon\_start=1  
/transl\_table=11  
/product="Glutamyl-tRNA Synthetase"  
/protein\_id="AAC68044.1"  
/db\_xref="GI:3328877"  
/translation="MTVQNVVRVAPSPGDPHVGTAYVALFNEVFARKYNGOMILRI  
EDTDQSRDDYANIFSAKWCIGRWDEPGVGGAYPPQSRERIEYKKAIELLQ  
TCAKCFRTPQLQENAVASLTGYRGYDRRYLYLSPPEVRQEEQQPQYIRLUK  
PLGSEVFEDQCGCVVPWADVDQVLKSGDGFYTFANVVDHLMGITVIRGEE  
WLSSTPKHLLYAFAGWEPQFPFHPMLLPDGSGLSKRKNPTSIFYRDAGYKEAF  
MNFLLMGYSMEGDEITYSMORLEAFDPKIRGSAVFDIRKLDWMKHYLNHEGSP  
ESLQELKGLWDEFLKILPLCQSRITTLADFVGLTSPFFFTAIPOYSKEELPSSL  
KQFQAVMLYSLVYLEKKDLWEKDFPYQGSKWLAFAQVHHKXAVIPLIYVAITGAK  
QGLPFDSEMLLQKTRARLTYAQNLLGGVSKVQVQVQDQALQDQPLEDFRLEP"  
complement (2721. .3272)  
/gene="euo"  
/note="synonym: CT446"  
complement (2721. .3272)  
/gene="euo"  
/codon\_start=1  
/transl\_table=11  
/product="CHLPS Euo Protein"

/protein\_id="AAC68045.1"  
/db\_xref="GI:3328878"  
/translation="MECLOQDTCVEAEQVQVQOQEEENAVPVTQSQRUSITOAAXLHNVT  
RQAIYVAIKQKUKASKTIRWEIDLOQLEDYRNRNYSRAKSTYQGEILLFNEKGFYSV  
QGVASMLDPEQKIYATRIGAMKGGRRGSAWVHVSEVDYRNDYLKKEAERKGS  
AAMREGFEALGADLADAEFNFS"  
complement (3677. .5431)  
/gene="recJ"  
/note="synonym: CT447"  
complement (3677. .5431)  
/gene="recJ"  
/codon\_start=1  
/transl\_table=11  
/product="ssDNA Exonuclease"  
/db\_xref="GI:3328879"  
/translation="MEKDSLSLQKQWDPYHNEVFLKKILKEFLHHPAATAQVLIISRG  
FQSIQIRDFLPQLSSLSHSTSLFDMKAVARLQAKANNEHMYGDDGVDGITGV  
TLIVEQLIQLGVKTSYCCSGTLFKQHGTEASLISOMQDGLSLITVDCGITTAKVEH  
AINKGIDIVITDHHMPTGKLPCHVAMLANPLDKNPYKNKELTGVAFLKVCATVEE  
LIQDASWKDKIDLLFLDLVSLGTTIADVGHLSGNNRILVSYGIIKIAKGRGLKLL  
CLSVDKSEVSTNLGIRITPKNLGSLRADSSQVKKLLSHDHPKNNIGTIVSELGSA  
NOERQIEAEVLRDVERILAAKPKLTAQSAI VLASPNWHSRVPIISARLARTNRPV  
AIALQDGIKGSRLTIGSPFLGLVKRCKESFFLSYCGHDFAAGLMIKEDQVSGFRKK  
FIHLVSSSLRKODAMRTSLDVMDFSRINFDLIASMELEPPFGKGNVSPVYTKAIO  
VAYPKLAGNHVKLYLNSGERNLEGTAFQCKDKISLLKANWNIPLDIATLIRMRSA  
RGAIRLLIQDFRIQIPRL"  
complement (5549. .9751)  
/gene="secD/secF"  
/note="synonym: CT448"  
complement (5549. .9751)  
/gene="secD/secF"  
/note="fusion"  
/codon\_start=1  
/transl\_table=11  
/product="Protein Export"  
/protein\_id="AAC68047.1"  
/db\_xref="GI:3328880"  
/translation="MKRNFKNLSIILVFLALYLVLPCLYYSRPLNRKIDGRAR  
KIIERTQVSDARDLTLURVAKILSTLKGDIKQHPISIPGVVNVHFLAEADALFI  
DNVHGEPSVPKSRARLYVVGSEFTGETVQISGLVSTLTKEDLSFVSYEDQVTV  
SOELENATNLFIESRPECGYSLWDSMPLEKIVYVTRNCSGLRLPARKQAPL  
SCLGTERDFTAPLARLEKALQHEEVEKARLEKDAASYLLHSRNLWNISIPKVGNV  
VDCSELSPCFSSMSLTADCKLPHFPAVANRQOLVGNERLDLELFAMEKQHIATR  
LNRSEGTGFGFVAVLRDEASGKIVLQHRVQRIVEHLTALVLRNPATETCDLSTE  
NPFVYREPLDGLGCFIPSPSRCHFSKGSVYIVFKGLRSIVAKYHAGAEAL  
LOHDLQNLVACFIHTDAVSGSLGEDRVLEIKPELQRIIRIWGEFVQSGKASLEVR  
VDRLARVNRIRIETQHAELVRWDEQVRQACSMNPRVLRRAIPHNKIIFFENLKLR  
KYSRGEHVLRFGTDFVGGKQIRIAFRDHQGLLTDKAGIDKYSDELAYRLNKLGVSEV  
GMQREGDHIQVSPGAAGISSADILGTSKWSFHVWNEQFSSRSPLEYVQTFDYLW  
TARSLDECSPOAIRNLAGALFHGDNGSAPANVRVAVEKLRBEAGLSFSKELEGSSALD  
TQSMIAIEKREQVNLUMIVFRNHALEGASLKNIRPEFAVGEVGLNFGVKKKATP  
SDGRETVPQGFHMTSKFCQGVSGTKNGLFGGRGWRMAVLDGVVIGDPVNLVPLK  
DHASVGNFSYREVHRLAADLKGAMSFIPETLSBEVSPSELGSSQVQVQILSVLGL  
VVLIVMSVYVYFGVGIASVLLNLLIWAQVLDAPLTLSGLAGIILAMQAVDA  
NVLVPERIEEYLLTSLSESVEAGYKKAISAFIDSNLTILASALLMLDTGPIKGF  
ALTLIGIPSSMFTALFMTKFFVFWQKTRTQLHMMNKFIGIKHNFLEKCRLLWV  
SGVVLVLCVGGFGAWDSVFGMDGGYALTDSDVEYNPEQMCVSRKRPQQIIGL  
SSRDYRVLRGDSSEKVIYSLQALDRVEIGESBQESQKGSYHLARLVQLVSDSS  
TTSMDVDSRGSVFGVSGLSNMRMTQAVIALFGALGIIIFYVSRFEWYAFSAICA  
LMHDLATCAVLVLFQRIQIDLOAGLMTVLGYSLNNTLIIFDRIREDREKL  
FTPMPLINDALQKTLGRTVMTATTLSVLVILLFVGGGSIENFAIMTIVGILLGTL  
SLIYAPPLLLFWVRKEEQNSLR"  
BASE COUNT 3118 a 2399 c 1722 g 2750 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1 6e-166 Length: 9989  
Score: 1937.00 Matches: 390  
Percent Similarity: 78.68% Conservatives: 64  
Best Local Similarity: 67.59% Mismatches: 73  
Query Match: 69.73% Indels: 50  
DB: 1 Gaps: 5





|   |  |             |             |
|---|--|-------------|-------------|
| TITLE<br>JOURNAL<br>COMMENT<br>FEATURES<br>source | Direct Submission<br>Submitted (01-MAR-2000) The Institute for Genomic Research, 9712<br>Medical Center Dr, Rockville, MD 20850, USA<br>On Jun 1, 2000 this sequence version replaced gi:7190754.<br>Location/Qualifiers<br>1. .12442<br>/organism="Chlamydia muridarum"<br>/mol_type="genomic DNA"<br>/db_xref="taxon:83560"<br>complement(86. .544)<br>/genes="srp"<br>/locus tag="TC0726"<br>complement(86. .544)<br>/genes="srp"<br>/locus tag="TC0726"<br>/notes="similar to SP:P18587; identified by sequence<br>similarity; putative"<br>/codon_start=1<br>/transl_table=11<br>/product="sulfur rich protein"<br>/protein_id="AAF39536.1"<br>/db_xref="GI:7190755"<br>/translation="MSTPIVSGVTSONNSSENVNARSILTKERASKILSSTAFKVV<br>GLAVGIFLIVLSVILFIPATASNPYLAIPAILGCYNIIGILSMNKGCSSEAK<br>WKLCKNVLKTSIEDILDDGELNNSNKIFTDDLSRVEDIVITLSRRNSVA"<br>complement(727. .2391)<br>/locus tag="TC0727"<br>complement(727. .2391)<br>/locus tag="TC0727"<br>/notes="similar to GB:X00364, GB:J00120, GB:M13930,<br>GB:M13930, GB:X00247, GB:X01903, GB:X00535,<br>GB:M13930, GB:X01903, GB:X01904, GB:X01906, GB:X01907,<br>GB:X00057, GB:X01905, GB:X01906, GB:X00559, GB:X01906,<br>GB:X00057, GB:X00558, GB:X66258, GB:X73256, GB:X02277,<br>GB:X02280, SP:P01106, PID:1196446, PID:188967, PID:188975,<br>PID:188977, PID:312410, PID:515633, GB:X00364, GB:J00120,<br>GB:M13930, GB:M13930, GB:X01903, GB:X00247, GB:X01903,<br>GB:X00535, GB:X01905, GB:X02275, GB:X01904, GB:X00196,<br>GB:X00197, GB:X00057, GB:X00058, GB:M13929, GB:X00559,<br>GB:X01906, GB:X02276, GB:X00558, GB:X66258, GB:X73256,<br>GB:X02277, GB:X02280, SP:P01106, PID:1196446, PID:188967,<br>PID:188975, PID:188977, PID:312410, and PID:515633;<br>identified by sequence similarity; putative"<br>/codon_start=1<br>/transl_table=11<br>/product="60 kDa outer membrane protein"<br>/protein_id="AAF39537.1"<br>/db_xref="GI:7190756"<br>/translation="MRIGDPMNKLIRRAVTIFAVTSVASLFSAGVLETSMAESLSTNV<br>ISLADTKAKETTSQKDRKKNHRTSVVRKEVTAVRTDKAVEPRQSCFGKVTYV<br>KVNDRNVEIVQSVPEYATVGSVPLEITAIKRDVVDVITIQOLPCEAEFVSSDPAT<br>TPTADGLVWKIDRLGGESKLTVMVKPLKEGCECTAATVACPIRSVTKGQPAI<br>CVKQGESACLPVTYRNVNQGTATARNVVENPVDGYAHASGQRLVLTIGD<br>MQPGORTIVFCPLKRGRTNIAVSYCGHKNASVTTVINEPCVQNVIEGADMS<br>YVCQYEVYISVNPGLDVLVDVIEDTLSPGTVVEAAGAIQISCKNLVLELNP<br>ESIQKVLVRAQTPGQFTNNVVKSCDCGICTSCAEATTYMKGAATHMVCVDTDCP<br>ICVGTNTRVICTNVRGSAEDTNVLLKFSKELQPISESGPTKGTITGNTVVFDSL<br>RLSGTVERSVTLKAVSAGDARGEALLSDTLTPVSDTENTH1Y"<br>complement(2539. .2805)<br>/genes="omp3B"<br>/locus tag="TC0728"<br>complement(2539. .2805)<br>/genes="omp3B"<br>/locus tag="TC0728"<br>/notes="similar to GB:U03644, and PID:476105; identified<br>by sequence similarity; putative"<br>/codon_start=1<br>/transl_table=11<br>/product="cysteine rich outer membrane protein 3"<br>/protein_id="AAF39538.1"<br>/db_xref="GI:7190757"<br>/translation="MKRTALLAALCVSVLSLSSCCRIVDCCFDPCAPIKSPCESKKR<br>EVNGGNSCNGYVPSFCGGELDHETKQGPQARGIQADGRCRQ" | gene<br>CDS | gene<br>CDS |
|   | 3133. .3357<br>/locus tag="TC0729"<br>3133. .3357<br>/locus tag="TC0729"<br>/notes="similar to GB:L07795, and PID:148862; identified<br>by sequence similarity; putative"<br>/codon_start=1<br>/transl_table=11<br>/product="conserved hypothetical protein"<br>/protein_id="AAF39539.1"<br>/db_xref="GI:7190758"<br>/translation="MKRFFFLMIATAALLTQCGSYEPKREYILQRCSPFLFKHAQAI<br>NFSLAQAVEKFTVITVQPSLNLKELSON"<br>complement(3354. .4874)<br>/genes="gltx"<br>/locus tag="TC0730"<br>complement(3354. .4874)<br>/genes="gltx"<br>/locus tag="TC0730"<br>/notes="similar to GB:U03554; identified by sequence<br>similarity; putative"<br>/codon_start=1<br>/transl_table=11<br>/product="glutaryl-tRNA synthetase"<br>/protein_id="AAF39540.1"<br>/db_xref="GI:7190759"<br>/translation="MTFQNVRVVAPSPGTDPHVGTAYMALNEVFARKFNKGMILRI<br>EDTDQTRSDDEKNIIFSALKWCGIRWDEPDIGGPGYRQSERTEIYKKAIEILQ<br>TITAYKCFATPOELQEMVAVASTGLRGYGRYRYLSAEVRRQEQOQPYTIRLUK<br>PDYSGVFEDQCKGRVVFADVDQVLVKSDFPTTFYFANVVDDHLMGITVLRGEE<br>WLSSTPKHLLLYAEFGWEPQFPHMLPDLNPDGSKLSKRNKTSIFYFNKDYKKGAF<br>MNFLLTUMGYSMEGDEIYSMORLIEAFDPKRTGRSGAVFDIRKLDMMKHYLNHGGSP<br>ESLLKELKDLANDDEFLLKILPLCOTRITTLADFIETLFFFMALPEYSKEDLLPSL<br>GHCQAAVMYLSLVKLEKDLWEKFFPYGSGKWLAEAFQVHHKAVIPLLYVTITGEK<br>QGLPFDPSMELIGKARTRALTAQNLGCVPKKQQQIDKALQDQPLBEIRPLDF"<br>complement(5124. .5681)<br>/genes="euo"<br>/locus tag="TC0731"<br>complement(5124. .5681)<br>/genes="euo"<br>/locus tag="TC0731"<br>/notes="similar to GB:V00497, GE:M25113, GB:U01317,<br>GB:M36640, GB:M25660, GB:A01592, GB:A18171, GB:M11428,<br>GB:M25079, GB:V00500, GB:L48213, GB:U0223, GB:V00499,<br>GB:V00498, SP:P02023, SP:P02042, PID:1066755, PID:1066758,<br>PID:1066761, PID:1066764, PID:1066767, PID:1066770,<br>PID:1066773, PID:1066776, PID:1066779, and PID:455998;<br>identified by sequence similarity; putative"<br>/codon_start=1<br>/transl_table=11<br>/product="euo protein"<br>/protein_id="AAF39541.1"<br>/db_xref="GI:7190760"<br>/translation="MECQQDTWAEVQVEQVEQEEENAVSANQRVSTOAKLHN<br>TRQATYVAKKKLAKSKRTWEIDLOLEIYRRNRYSAKTSYQELLFDNEKGFYS<br>VQGVASMLNPEQKIYVATRICAMGERRGSAWVHVSEVDYRNDYLLKEAERKGN<br>LSAMEGFTLCASELLADEDFVS"<br>complement(6143. .7891)<br>/genes="recJ"<br>/locus tag="TC0732"<br>complement(6143. .7891)<br>/genes="recJ"<br>/locus tag="TC0732"<br>/notes="identified by match to PFAM protein family HMM<br>PF03020"<br>/codon_start=1<br>/transl_table=11<br>/product="single-stranded-DNA-specific exonuclease RecJ"<br>/protein_id="AAF39542.1"<br>/db_xref="GI:7190761"<br>/translation="MEKDKLSALGQKWTYPFHNEBFLKIKIKEYLHPVIAQVLVSRG<br>FQSVQEIYRDLFPYQSPFLDPASSFLDMKAVARLLKSKENNEHVMYIGDVGITGV<br>TLLVFLQMLGVKTSYCCSGTLFKQGEITLSIQMLQDQVSLITVDCGITAQREVR   | gene<br>CDS | gene<br>CDS |

SINKQIDVIVTDHHTPTGKLPHCIAIATNPCLDKNSYPNKELTGVGVAFKLVCAATYEE  
 LKODASWKIDILHFLDVLSTGIADVRLSGENRILVSYKEITKGRKLGKLL  
 LKSGVDKSDINLGRITPKLNSLGRADSDSQVKLLLSQDPRVTGKIVSELAIV  
 NOERQTEAEVLRDVERITAPLANPKLTAQSAIVLASPVHRSRVIPIISARLARTYKPV  
 AIALQDICKGSRITGAPLANPKLTAQSAIVLASPVHRSRVIPIISARLARTYKPV  
 FIHLVNSLRKDEARTILPDLVCMDFSOISODLSSMELLEPPFGKGNPSVFFYTRAIQ  
 VRYPKLUSGNHVKULYNSGERNLEGTAFQGGQKISLIKANNMPLDIATYLRVMKRSV  
 HGAVRLIQDFRIQAP"  
 complement (8009..12211)  
 /locus\_tag="TC0733"  
 /locus\_tag="TC0733"  
 complement (8009..12211)  
 /product="secDF protein, putative"  
 /protein\_id="AAF39543.1"  
 /db\_xref="GI:7190762"  
 /translation="MKRNFRKNSLVIIFVAFALALYHLVPTCLYYSRPLNKKIEGREAH  
 KIIRLTOQVADARNDLTVRKRLSTLKLKRGSKPHLSIPGVNVHFPPLAEDAAVPI  
 DNVVHGHSVPTKARLYVLGSSSLTGETIIVOVSGSLVTSLTDEDFSVSVEKDLVV  
 SOELNTATNLPVNVESCDHGYSSLDWSPIEKVAHARGICSLRLLPTEKTOAEL  
 SRFGTQDFVAFUSRVENALHEDVKEYREELNEAAYLLHNRNLNRWNSPRVVGNV  
 VDCSELPFCSSMSLTQDGLLPHFDADIAKROQLVGDRLDLERLFAVEKQHIATR  
 LNRVETESGFARLDEAASGKIQLQHRVCORIVEHLTALVLRNPIAETCDLSIRE  
 NFPVSRPEMNTDLGCFIFSPDSSCRHFSKGVYIVFKGLRSIIAKCESSGPOEAGL  
 LQODLONLYACFIHTDAVMSLSLGRVLEIKPELQVIOVWGEFFVQSGEASLEVRD  
 VRDLAVVNRLEKQAQHAELVNDQYRQAQCSNMNPQVRLRAAIPIHKNVVFENFKLNV

## Alignment Scores:

Pred. No.: 3.31e-165 Length: 12442  
 Score: 1924.00 Matches: 384  
 Percent Similarity: 79.30% Conservative: 72  
 Best Local Similarity: 66.78% Mismatches: 73  
 Query Match: 69.26% Indels: 46  
 DB: 1 Gaps: 5

US-09-868-987-14 (1-552) x AE002341 (1-12442)

Qy 1 MetValSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20  
 Db 9679 GTAATTAGTATCCCGTGTAAATGTTCCGCTCAAGATCATGCAAGTGTTCCTGGGAAT 9620  
 Qy 21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40  
 Db 9619 TTTTCGTACAGGAGTCAATCGTTAGCTTCTGATCTTAAATCTGGGGCTATGCTCTTT 9560  
 Qy 41 ValProGluValLeuSerGluThrIleSerSerAspLeuGlyLysLysGlnCysThr 60  
 Db 9559 ATCCCTGAAATTTTAAAGTGAGGAAGTAGTTTCTCCAGAGTTAGCACACTCTCAGAGGTG 9500  
 Qy 61 GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80  
 Db 9499 CAAGGGATTCTCTCTGATTTTAAAGTACGTAGTGGTCTCTCATTTGCAATTAATGAGTGTAT 9440  
 Qy 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrp 100  
 Db 9439 TACAAGTTGGGGTGTCTCTGCTCGGTAGCCGCTCTTCTTAATTACTACTTATCTCGG 9380  
 Qy 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120  
 Db 9379 GCATCTATGCAGTATTAGATGCTCTCTTACTCTATCAGGACTAGCTGGAAATATCTTA 9320  
 Qy 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPhe 140  
 Db 9319 GCAATGGGGATGGCGGTGATGCTAAACGTCCTGGTCTTAAAGGATTCGAGAGAATAT 9260  
 Qy 141 LeuLeuSerGlnSerLysLysSerValGlnLysGlyTyrThrLysAlaPheGlyAla 160  
 Db 9259 TTATTGAGTAGAAGTTATACAGTCTGTAGAGCTGTGTTAAGAGAGCTTTTAGTGCA 9200  
 Qy 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180

Db 9199 ATTTTTCATTCTAATTAAACAATATTTTGGCTTCGCAATGCTTCTCATGTTGGATACA 9140  
 Qy 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200  
 Db 9139 GGACCAATTAAGAGGTTTGTCTTGGACCTCATCGGAATTTTTTCTTCTATGTTTCA 9080  
 Qy 201 AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisThrGln 220  
 Db 9079 TCITTTGTTTCATGACGAAGTTCTTTTCTTCTGATTTGGATACAAAAACAGAGACCCNA 9020  
 Qy 221 LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys 240  
 Db 9019 TTACATATGATGAACAAGTTTCATAGGGGTCAACATATAATTTCTTGAAGAATGTGAAGCGT 8960  
 Qy 241 LeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla 260  
 Db 8959 TTGTGGGTGTTTCCCGGGGTGTAGTGTGTTTATAGTTGTATTGGGCTTAGGCTTTGGTGCC 8900  
 Qy 261 TrpAsnSerValLeuGlyMetAspPheLysGlyLysValTyrAlaPheThrPheAsnProLys 280  
 Db 8899 TGGGACTCCATTTTAGGTATGGATTTTCAAGGGAGGATACCGCTGACCTTAGAT----- 8846  
 Qy 281 GluHisGlyIleSerAspValAla-----GlnMetArgGlyLysValVal 295  
 Db 8845 -----TCTGATGTTTGTGCCTATAATAATCCACATCAGATGTCTTCTGCTTTAAAA 8798  
 Qy 296 HisLysLeuGlnGluAlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySer 315  
 Db 8797 AAACGCTTCGCGGAAATTTGGGTTGTTCTCTCGTATTATCTGCTGTAAGCTGTAGTAGC 8738  
 Qy 316 SerGluLysIleLysIleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArg 335  
 Db 8737 TCTGGAAGAGTTAAGATCTATTTTCTCAAAATGCTTTAGCTCGTGTGACCAAGCCAAA 8678  
 Qy 336 -----AlaSerLeuLeuLysLeuThr 342  
 Db 8677 GGGGGGTATTCCGAACAGAATGGGCGGACCATCTAGCTCAAGTTTTCGAAGTATTA 8618  
 Qy 343 IleMetSerTrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeuTyrGly 362  
 Db 8617 TCCGATTCT-----AGT 8606  
 Qy 363 AsnSerLysArgAsnAlaLysPhe-----TrpSerLysValSerSer 376  
 Db 8605 GATGCTACTTCTCTCGCAGCATTTGAATCTCTCGAGGAAGTTGTTTAAAGTTAAGTTGA 8546  
 Qy 377 LysLeuSerLysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIle 396  
 Db 8545 CAGCTTTTCGAATAAAATGCCAACACAGGCAATCATGGCAATTATTTGGAGCTTTTAGGAATT 8486  
 Qy 397 IleLeuLeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAla 416  
 Db 8485 ATCTTACTTTATGTTAGCTTTCGTTTGAATGGAGATATGCTTTTCAGTGCATTTTGTGCT 8426  
 Qy 417 LeuIleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPhePheLeuLys 436  
 Db 8425 TTAATGCAATGATTTTGTAGGACCTGCTCTGTTATAGTCTCTGCTCTGCTCTTCTTTTCAT 8366  
 Qy 437 LysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeu 456  
 Db 8365 AGAATACAGATTGCTTACAGCAGTGGGGCGCATTTGATACGGTGTAGGCTACTCTTTT 8306  
 Qy 457 AsnAsnThrLeuIlePheAspArgIleArgGluAspArgGlnAlaAsnLeuPheThr 476  
 Db 8305 AATAATACGTTGATTTATTTTGTATCGTATTTCGAGAGACCGAGCTGAGAAATTTGTTCCACC 8246  
 Qy 477 ProMetHisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThr 496  
 Db 8245 CCAATGCCCATATGATCAATGATCTTTTCAGAGACTCTGGACGCTAGCGTCTATGACA 8186  
 Qy 497 ThrAlaThrThrLeuSerValLeuLeuMetLeuLeuPheIleGlyGlySerSerValPhe 516



|                   |  |
|-------------------|--|
|                   | VLIIRIAQDGGENAEQSAITLVRCGELEDKDFRVEVVGPAVSGDLTVTSTIGVLAA   |
|                   | AATVIYIWRFEWFGALGVAISMVDHVVFTIGFVFLGIENFNLTSAAITLTIIGYSIN  |
|                   | DTVTIVIKRRERYKKPMISMTIDVSLNQTLSTRTLTGLTWLLALLLYLFPGGVRI  |
|                   | RSFTFAMLFGVIGVFSSVITAAPVLIAFAKLRPDSKDAEDEKDDKENNTGAIGGPKA  |
|                   | V"   |
| gene              | complement (5493. .5840)<br>/gene="yajc"<br>/note="synonym: Atui563"<br>complement (5493. .5840)<br>/gene="yajc"<br>/note="identified by sequence similarity; putative; ORF<br>located using Blastx/Glimmer"<br>/codon_start=1<br>/transl_table=1<br>/product="preprotein tranlocase protein"<br>/protein_id="AAL42567.1"<br>/db_xref="GI:17739991"<br>/translation="WFESQAFADASCAGSFGSGLEMLFLFAPLMVVMVFFILIRPQR<br>AQMKRQETLSIRRGDOVLGGIGVKTKVIDDNELEVEIAEGVKIRARAARYIAEAV<br>RVKGEEVKTAAAS"<br>6048. .6920<br>/gene="Atui564"<br>6048. .6920<br>/gene="Atui564"<br>/note="identified by sequence similarity; putative; ORF<br>located using Blastx/Glimmer"<br>/codon_start=1<br>/transl_table=1<br>/product="conserved hypothetical protein"<br>/protein_id="AAL42568.1"<br>/db_xref="GI:17739992"<br>/translation="MIREETAULLSELRLTNNAVERLAGPAANDWNAAACFVWAP<br>LNSFLQPVRNRITALKIRGDVHDVDRDIHLHTLRFAEGFAANNVLLMGARGMKGSLL<br>VKAVEDVRAASGVSLKIVEVHRDIIHTLPALLDIKISDVRLFCDDLSPDHDTTA<br>YKSLLKAALDGIEGRDPONLVFYATSNERHLLPHRMENEQSTAINSEAVEEKVSLSD<br>RFGVLGHFKKSCQSDYLTMIDSYAHEFDGLUDRTLHLEALEWATTGRGSRGVAVQY<br>IODAAGRLLRKPLDR"<br>complement (6926. .7555)<br>/gene="cysE"<br>/note="synonym: Atui565"<br>complement (6926. .7555)<br>/gene="cysE"<br>/note="identified by sequence similarity; putative; ORF<br>located using Blastx/Glimmer"<br>/codon_start=1<br>/transl_table=1<br>/product="serine acetyltransferase"<br>/protein_id="AAL42569.1"<br>/db_xref="GI:17739993"<br>/translation="MHCVFLEFNMSVFVKIYFALSAPIVAPHIFLLMSSRHNDPIFRD<br>IEVNSEKLLGKKIGSRGLVASVVELLSLAPEPNLFYRYVPARFPFKFLSRPENTLH<br>IHTRITIGPGFIIOHGFTATIASGISGADCMINOQTGYVDSTCAPKIGDNVTINAGA<br>KVLGNIVIGDMAKYGANAVNYNTPADCVVVGVPARIIRAGARVDEAL"<br>complement (7578. .7763)<br>/gene="Atui566"<br>complement (7578. .7763)<br>/gene="Atui566"<br>/note="putative; ORF located using Glimmer"<br>/codon_start=1<br>/transl_table=1<br>/product="hypothetical protein"<br>/protein_id="AAL42570.1"<br>/db_xref="GI:17739994"<br>/translation="WSFGNTRRVVATRNPITYSFFTQQTKAFMIICLDWRNRLPAG<br>ITQSSEAWEBWRFLFA"<br>8013. .9209 |
| gene              |  |
| mmnt Scores:      |  |
| No. :             | 4.06e-64 Length: 10046   |
| e.                | 817.50 Matches: 187  |
| ent Similarity:   | 58.42% Conservative: 139   |
| Local Similarity: | 33.51% Mismatches: 190   |
| Match:            | 29.43% Indels: 43  |

```

Db      3484 ATCA----- 3481
QY      356 gProArgPheLeuTyrGlyAsnSerLysArgAsn-----AlaLysPheTrpSerLysVa 374
Db      3480 ----CGCTGGTTCGGCGGCACTTGAGGACAAATACGACTTCCTCGCGTGGAGTC-GT 3426
QY      374 lSerSerLysLeuSerLysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLe 394
Db      3425 TGGTCGGCGCTCCGGTGACCTACCGCTACTTCCAGCATCGGTGCTCGCTCGCAT 3366
QY      394 uAlaIleIleLeuLeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaVa 414
Db      3365 GCGCGCATCATGCTCATCTCGTTCGCTTCGAGTGGCAGTTGCACTCGGTGCGGT 3306
QY      414 lCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPhePh 434
Db      3305 CATCTCGATGGTGCATGACGTGCTTCCACC-----ATCGGCCTTTTCGT 3261
QY      434 eLeuLysLysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTY 454
Db      3260 CTTCCTGGGTATAGATTCAACCTGACCATGATCGCGCGATCTCGAGCATCATCGGTTA 3201
QY      454 rSerLeuAsnThrLeuIleIlePheAspArgIleArgGluAspArgGlnAlaAsnLe 474
Db      3200 TTCGCTGAACGACACTGTCGTCATCTACGACCGTGTCCGTGAAAAACCTCCGCGCATCAA 3141
QY      474 uPheThrProMetHisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrVa 494
Db      3140 GAAGATCGCGCTTCGATGATCATCGCATGTCGTGAACACGACCCCTGTCGCGCACGAT 3081
QY      494 lMetThrThrAlaThrLeuSerValLeuLeuMetLeuLeuPheIleGlyGlySerSe 514
Db      3080 CCTTACCGCGCTGACCGGTCTTCGGCACTGCTGCGCATGTATCTCTCGGTGCGGAGGT 3021
QY      514 rValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLe 534
Db      3020 CATCGTCTCTTCACTTCGCGCATGTTGTCGCGGTGCGTATCGCGGTGTTCTCTCGGT 2961
QY      534 uTyrIleAlaProProLeuLeuLeu---PheMetValArgLysGluAsnArg 550
Db      2960 CTATATCGCGCTCGGTGCTGATCGCTTCAAGTCGCTCGCTCACAGCAG 2909

RESULT 9
LOCUS   AE008079/c 10066 bp DNA linear BCT 18-DEC-2001
DEFINITION Agrobacterium tumefaciens str. C58 circular chromosome, section 137 of 254 of the complete sequence.
ACCESSION AE008079 AE007869
VERSION   AE008079.1 GI:15156644
KEYWORDS
SOURCE    Agrobacterium tumefaciens str. C58 (Cereon)
ORGANISM  Agrobacterium tumefaciens str. C58 (Cereon)
REFERENCE 1 (bases 1 to 10066)
AUTHORS   Hinkle,G., Slater,S.C. and Goodner,B.
TITLE     Complete Genome Sequence of Agrobacterium tumefaciens C58 (Rhizobium radiobacter C58), the Causative Agent of Crown Gall Disease in Plants
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 10066)
AUTHORS   Hinkle,G., Slater,S.C. and Goodner,B.
TITLE     Direct Submission
JOURNAL   Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street, Cambridge, MA 02139, USA
FEATURES
     source
         1..10066
             /organism="Agrobacterium tumefaciens str. C58 (Cereon)"
             /mol_type="genomic DNA"
             /strain="C58"
             /db_xref="taxon:181661"
         67..1002
             gene

```

```

/gene="AGR_C_2870"
67..1002
/gene="AGR_C_2870"
/notes="hypothetical protein PA0239 (imported) - Pseudomonas aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_C_2870p"
/protein_id="AAK87343.1"
/db_xref="GI:15156645"
/translation="MTSQTFNDIAKGMTINAGCMLVLPVMDIAIKYMAVSVGMSPAQ
VTEYRFVFPFLPLLVSPGALFSAKRPMNCLRGVLHAGASLMFFAAVYKMPILA
DVFAIYFVFPFLMMSALFLGDRVGKRWTAIVCGFGGALIVIQSFEFFGTSLLP
VACAFYTLVFLNRAIGDASPLVMQTMGTGLFMAAALFAGDAGDKDFAVSLP
QSAFALVLLVGLSGYMHLLIVAFRLAPISLLAPFQYFEILSITLIGYALFDFP
TPSNWGLIIVASGDFIWRERQQPASTPPELGG"
complement(1014..1544)
/gene="AGR_C_2872"
complement(1014..1544)
/gene="AGR_C_2872"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_2872p"
/protein_id="AAK87344.1"
/db_xref="GI:15156646"
/translation="MNPDRVVRVSSRESKGFGTGEGSRMNLAIWIVFVLLVAVVA
WRMSRATEDSSHDVQAIIMDFARAPFSEAIRSLHMTADGGAAVRLHDKKAGFMNM
GSHYVCMDIPERIRVESLETGDFVYTFDMPKYSYRFRKSAGEAAEVSLLGLSL
VEAQDHDEGPIPSGV"
complement(1596..2441)
/gene="AGR_C_2873"
complement(1596..2441)
/gene="AGR_C_2873"
/notes="(AC007190) F23N19.9"
/codon_start=1
/transl_table=11
/product="AGR_C_2873p"
/protein_id="AAK87345.1"
/db_xref="GI:15156647"
/translation="MEPTENLDVCLATLDRDRYLACLSPADKRNALAAALYAFNA
ETARIDSVREALPGEVRQWNRDLVDGNPHGDSQSHPVAAALLTIEHYRLPRPVLG
NMIEARIFDLVDLFDNRNAGYAGETASALIQLASLSAEDAHAASAAAGHAGVA
QAMAGMLLPLHRRRGQYIPADMLAAGLDRFTFLAGDQRQIGIAGVEIFAHAHALD
HLEKAPRATISKNLFAALPVALSGPVIAAARKAGAGVFDGLQLSLRQWLLAKAS
FLKRP"
complement(2465..2857)
/gene="AGR_C_2875"
complement(2465..2857)
/gene="AGR_C_2875"
/notes="(AF157831) hypothetical outer membrane protein"
/codon_start=1
/transl_table=11
/product="AGR_C_2875p"
/protein_id="AAK87346.1"
/db_xref="GI:15156648"
/translation="MAGGIEIRPAHPGPRAPLDAYGNGGFRFADMSHRGSLLLPSGI
YGEPIDAKELTVEHFERVLAQAQDIEVLLIGTGDGMVLPKELRAAFKAGISIDPM
STGAAVRYNIILSSRAVAAALIAVEG"
complement(2857..5487)
/gene="AGR_C_2877"
complement(2857..5487)
/gene="AGR_C_2877"
/notes="probable protein-export membrane protein - Deinococcus radiodurans (strain R1)"
/codon_start=1
/transl_table=11
/product="AGR_C_2877p"
/protein_id="AAK87347.1"
/db_xref="GI:15156649"
/translation="NRILTGCATVATPTLTETEMLHFGRWKTVFIWLLVLAASFVFIAS
NLFSEKQLESMPAWYKDKNKTGLDGGSHIMLKIERSDIVKERLETIVGQVDTQLR
DANIRYSLGTGNGQIQIVRISDPAQMEAAKTALRDLTQVPSAGTLVGGSITEVTMNDG

```







```
KKIRAEKDQKQRFVSGAVERGVSKPQADFIIDLAKFADYGFNKSAAAVAVSVYQT
AYLKAHYVEFLAASTMDMGNTDKLADROQDALRGLIEVLAPSVMTSFRPFYGENR
IYYSMAALUKGVGDAVEHIVARGEKPFKNLADFCERVDPKIVGRKVFESILMAGALD
CFGHDAACOMAGVERMGLASLAQONAVSGOAGDFGASLGSQSQALNLPATDPMLAAD
RLHREFQVGVFSLSHDDEYKAAALQKMRVONWAEFFSAVRKRGASAGLACTVTSKQB
RKRTIRGNKGVVAFSDTSQGEYAEALVFSALAOYRDLLEAGRSVVITVAEEDRPEGVNL
RINSVQSLEESARIQKALRIFVRNEGPAISIIQSOLTQRGESQVSIIIVKEAQGEVE
IALOGGYRVSQIASAMRAVPGVVEVELV"
/gene="ml10871"
/complement (4811..6823)
/complement (4811..6823)
/gene="ml10871"
/codon_start=1
/transl_table=11
/product="propionyl-CoA carboxylase alpha chain precursor"
/protein_id="BAB48367.1"
/db_xref="GI:14021756"
/translation="MFTKILLIANRGEIACRVIKTARKMGIAIVAVSDADRAVHEM
ADEAVHIGPSPAAQSYLPEKIIIACKETGAQVHPGYGFLSERAACEALEKEGIVF
IGPKPAIKAMGDKIESKFAANAARVSTVPGLGVENADHAEKIAEIGYPMVIKAS
AGGGKGMRIANWQSEVRDGFDRARSEAKSSFGDQDVPIEKVVDPRIHIEIOVLADAH
GNALVGERECISIORNQVAEEAPSPFLDAKTRKAMGEQSVALARAVDQSGACTVEF
IYDKDKNYFLEMNTRLOVHPVTELVTDGLVOMIRKVAAGEKLSUKQSDIKLNGWA
VESRUYADBPFLNPLPSIGRLTRYRPPPEQGFGDVVIRNDTGVTGEGSEISMFYDPLA
KLCTWAPTRIIEAIDAMSEVLDSFVVDGIEHNIPLAALMQHPRWREGAISTGFLAEY
PDGPAPVPSNREKRAVLASIIVAVELLRRDLRLGLGRLLAPHSGLVKRDVAVRIGDDY
LOASTLEGMISIPMEIDLSIDGKALTVSSDRPCDLIRWCTVGRKSVQAQIRVPANG
LRIAWKGSVTARAMPLPRIASLERLMPKVPADPTSKLLCCMPGLGVLSIATVEQGEVK
AGETUAVVEAKMENVLRAERDLVVVSKLNAPGDSLAVDAVIMEBFA"
7023..8150
/gene="mlr0872"
7023..8150
/gene="mlr0872"
/codon_start=1
/transl_table=11
/product="glutathione dependent formaldehyde
dehydrogenase"
/protein_id="BAB48368.1"
/db_xref="GI:14021757"
/translation="MKTRAAVAVAAGKLEIMEVDLQPREGEVLVRIKATGICHTDE
FTLGSADREGFPALPIHLEGAGVVDVGKGTYSVKGDHVIPLVTPBCRCQPSCLSRK
TNLCATIRATQOGLMPDQSGRSRFSKDKIIFHYGCSFTSFNFTVLPIEALAKVNPDR
FDKICYCGVTTGIGAVINTAKVEQGAATVAVGLGIGLNVIOQLKLAGADMIIIGVD
LNNDKAMRGFTGFHFVNPKIEDIGDIVPHLVNMTKRGADQIGGADYTFDCTGNTKYM
ROALEASHRGKSVVIGVAGAGOEISTRPQLATGRTWMTGAFGGARGRDVDRIVD
WYMDGKIQIDPMTLLKLEINDKGFDLMEHGKSIIRSVVY"
8113..8787
/gene="mlr0874"
8113..8787
/gene="mlr0874"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB48369.1"
/db_xref="GI:14021758"
/translation="WRASRSVASFTEASHTPTDTRPATSVVQNWEEKRAEKLHPRI
DNGLPKESAGGCTGACCTSKPVKVKQGLAHNHACGCTCKWKEGALFSVAVVA
GTGVDVTAFNDGKLKVDVADALIQRHACTGCGVHHGVPVRDHFAPGLTFIHPERFVE
DWSGPGFTFAVSSIIYESVDPKRMDGIRQAIRLTIGLEPYDCLNPLMDYIATWTAKX
SGALPA"
8818..9318
/gene="mlr0875"
8818..9318
/gene="mlr0875"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB48370.1"
/db_xref="GI:14021759"
/translation="MFPAPVRFGIEMPAFTIYVDADAPVKAEEVKEKVAERLGVVVTV
SNGLRPSRDMIRNVVSKGADAADWIVENAKSNVDVVTADIPLAARAVALLGAHV
GPTGRPFPTETIGMAVMRDLKQLHRETGESKGNAGFAPQDRSRFLGELDRILRRAL
KSATPG"
```

```
gene
CDS
9338..10027
/gene="mlr0876"
9338..10027
/gene="mlr0876"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB48371.1"
/db_xref="GI:14021760"
/translation="METAMAAETPIRPVQRHMLFAMSACFGALAFVALLHAPLAY
SIGANAPFAAYVILVVSQMPKFTSKYLSSKNARATDQVLVIFAVTLVVVIVAVVLSFL
LINQNRGNSVEIGFALLSIPLCWFTIHAMAAHYAHVYMMDGDADVAETFKKIPVGG
LLFPQDKRPEGWDFLYPSAVIGMTAOTADTNISTTHMRCVVLVHSILSFFFTVIVAA
AVNLAVSLGGP"
10125..11006
/gene="mlr0877"
10125..11006
/gene="mlr0877"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB48372.1"
/db_xref="GI:14021761"
/translation="MQSVARQSQATGDPAPVSSDSTLIYQSRWTRLTHMIWISLFF
MLLSGLQIFNARPOLYIGKSGFGYNNITFEIGAVNTDAGPRGTEIFGHRFDTTGLV
GWSGPRNETARGFPSWATIPSYVDLGTARVHFFFAWILTTLIVMLVASLVNGHIR
RDLAPRLDDLKRLUPRIDVDAHARLKFHTREYNTLQKWAYGGVLFVLLPLMIITGLAMS
PSMNSVLPLFNEVGGRGQTARTIHTFMLLLVGFFIHIILMILAAGFINEURSIIITGM
YRTDPPAHDDKKTTERSA"
10975..11817
/gene="mlr0879"
10975..11817
/gene="mlr0879"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB48373.1"
/db_xref="GI:14021762"
/translation="MTTRPRGVPDMAKFEISNFOISRRKFLTAASLGSAGIMLSGC
DAFDSQLSISGSLRSFLENANGUTYRQRLLAGRDALAPEFTEADIRQORPQNGVTAP
DDDTYKGLANNFADWRLEVSGLVERKPLSLRQOLQNPSTQIIRHDCVGEWSCIAK
WTDTPLSLDLDQAVRQYVYVNFHCLDTIDRSLGDIKYITGIDIDARHPQTIILAY
GLNGKPLPEVENGAPLVRVERQLGYKMPKLYKIELVDGFANVGGGRGYWEDNGYDM
YGGI"
complement (11866..12867)
/gene="ml10880"
complement (11866..12867)
/gene="ml10880"
/notes="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB48374.1"
/db_xref="GI:14021763"
/translation="MGELIMSAPVAGLTSKNRIAAEDVLMLRRDRFADGVVTVRGEAEA
LPALDQAQDKCAEWAPFVEAVTDVIVHQEKSGVVISQDNADWLVRTISRDGWDVSR"
Alignment Scores:
Pred. No.: 2,03e-57 Length: 349619
Score: 766.50 Matches: 183
Percent Similarity: 56.42% Conservative: 129
Best Local Similarity: 33.09% Mismatches: 204
Query Match: 27.59% Indels: 37
DB: 1 Gaps: 13
US-09-868-987-14 (1-552) x AP002996 (1-349619)
Qy 1 MetValSerSerProIleLeuAsnValPro---LeuLysAsnHisAlaSerValSerGly 19
Db 202374 GTGATCTCGCGCGCGCGAGATCCGTGAACCCATCTCTCGCGCGCACCGCCAGATCTCAGGC 202315
Qy 20 LysPheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSer 39
Db 202314 AATTTCACCCGCCAGAGCGCAATGATCTCGCGCGTCTTTTTCGAGCGGTGCGCTCCG 202255
```

QY 40 PheValProGluValLeuSerGluClnThrIleSerSerAspLeuGlyLysLysGlnCys 59  
Db 202254 CGGACGCTGACGGTATCGAGAACGCACTCTCGGTCCGGCCCTTGGCCAGATTCGATT 202195  
QY 60 ThrGlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerVal 79  
Db 202194 CATCGCGCAGGTGGCGGATCATCGTTGATCTCTCGTTCGTTTCATGTTCTGTC 202135  
QY 80 TyrTyrArgPheGlyValIleAlaSerGlyValLeuValLeuLeuLeuLeuLeu 99  
Db 202134 GCCTACGGCTCTCTCGTTCTCGCCAAATCGCGTGGCGGTGCACGTGGCATGATC 202075  
QY 100 TrpAlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleVal 119  
Db 202074 GTCGGCTGTTGTCTGCTCTGGGCGGACGCTGACCTTCCCGGTATTCGCGGTATCGTG 202015  
QY 120 LeuAlaMetGlyMetAlaValAlaAsnValLeuValPheGluArgIleAtrGluGlu 139  
Db 202014 CTGACCATCGGATGGCGGTGATTCACACGTGCTGATCTACGAACGCATCCGTGAAGAA 201955  
QY 140 PheLeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGly 159  
Db 201954 CGACGGGTGGCGCTCGGTATCCAGCGCATCCAGCACCGGCTTCTCGAAGCGCTGGCG 201895  
QY 160 AlaIlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPheLeuAsp 179  
Db 201894 ACCATGCTGATTCACAGTTACGTGCTGATCCGCCACCGGTGGTGTCTATCTCGGC 201835  
QY 180 ThrGlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPhe 199  
Db 201834 ACCGCCCGGTGAAGGCTTTGCCATCACCTACGCCATCGTATCTTGACACCGTGTTC 201775  
QY 200 ThrAlaLeuPheMetThrLysPhePhePheMetLeuTrpMetAsnLysThrGlnHisThr 219  
Db 201774 ACCGCTTTCACCTTCACGGCGCTGCTGTGTCGATCTGGTCCGCCGCGCTCCGAAG 201715  
QY 220 GlnLeuHisMetMetAsn---LysPheVal-----GlyIleLysHisAspPheLeuArg 236  
Db 201714 GAATTCGCAAGGCGCGGTGACCTTCATTCGCGGGGACCAAGATCCCTTCATG--- 201659  
QY 237 GlyCysLysLysLeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeu 256  
Db 201657 GGTATCCGCCGC---TGAAGCTTCGCGTGTGCGATTCCTGTCGGTCTCTCCGTGGTG 201601  
QY 257 GlyPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGlyTyrAlaPheThr 276  
Db 201600 CTGTTTCATGACGTCGACATCAATTACGGCATCTCAAGTTCAGGCGGCTCGATCGAG 201541  
QY 277 PheAsnProLysGluHisGlyLeuSerAspValAlaGlnMetArgGlyLysValValHis 296  
Db 201540 GTGAAGTCCAAGAC---GGCGACGCCAATCTCGCGATATCCGCGGCGGCTGATGGAG 201484  
QY 297 LysLeuGlnGluAlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSer 316  
Db 201483 -----CTCAACATCGCGCAAGTGCAGGTGCAGAGTTCGCGCGCGC 201442  
QY 317 GluLysIleLysIleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAla 336  
Db 201441 -----AACGACGTGTCGTCGCGCTCGCGCACGAGGATGGCGC 201403  
QY 337 SerLeuLeuLysLeuThrIleMetSerTrpArgTyrCysGlyIleValValArgAsnArg 356  
Db 201402 GAGAAATCCGAGACGCGTATCGAC-----AAGTACCGCGCGGAG 201361  
QY 357 ProArgPheLeuTyrGlyAsnSerLysArgAsnAlaLysPheTrpSerLysValSerSer 376  
Db 201360 TTGCAGGACCAAGTAT---GATTCGCGCGGTCTCAAGTCTGGGCGGCGGCTTCGCGC 201304  
QY 377 LysLeuSerLysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIle 396  
Db 201303 GAATGCGCAAG-----CAAGGCACCATCGCCATGCTGTGTCGCGCTGGTTGGC 201256

QY 397 IleLeuLeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAla 416  
Db 201255 ATCTCGTCTATGCTCTGTTTCGAGTGGCAGTTTGGGTCCGGCCCATCGTTGGC 201196  
QY 417 LeuIleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPheLeuLys 436  
Db 201195 ACAGTGCAGATGTGTCATGACG-----CTCGGTTTCTTCGTCATCACC 201151  
QY 437 LysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeu 456  
Db 201150 GGCTCTGAATTCACACAGTCTGCTGGCGGCGATCCTGACCATCATCGGCTATTCGCTG 201091  
QY 457 AsnAsnThrLeuIleIlePheAspArgIleArgGluAspArgGlnAlaAsnLeuPheThr 476  
Db 201090 AACGACCATGCTGTGCTATGACCGCTCCGAGAGATCTCGAAAAATACAGCGAATG 201031  
QY 477 ProMetHisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThr 496  
Db 201030 CGGCTCCGCGAGCTGTGTGAACAACGCGCATCAACGAGACGCTGTGCGGAACGAGCTGACC 200971  
QY 497 ThrAlaThrThrLeuSerValLeuLeuMetLeuLeuPheIleGlySerSerValPhe 516  
Db 200970 TCGTGACACGAGTCTGCGCTGTCTGCGCGCTGTCTGTCGCGCGAAGTCACTCCGC 200911  
QY 517 AsnPheAlaPheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIle 536  
Db 200910 TCTTTCACCATGGCAATGCTGTTCGCGGTGTCTGCTACCTATCTTCGATCTTCATC 200851  
QY 537 AlaProProLeuLeu---LeuPheMetValArgLysGlu 548  
Db 200850 GCGGCCGCTGCTGATCTCTGTTCAAGCTCGCGCGCAG 200812  
RESULT 11  
AP005952/c 300600 bp DNA linear BCT 28-MAR-2003  
LOCUS Bradyrhizobium japonicum USDA 110 DNA, complete genome, section 18/31.  
DEFINITION AP005952.1 GI:27352868  
ACCESSION AP005952 BAO000040  
VERSION AP005952.1  
KEYWORDS Bradyrhizobium japonicum USDA 110  
SOURCE Bradyrhizobium japonicum USDA 110  
ORGANISM Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
Bradyrhizobiaceae; Bradyrhizobium.  
REFERENCE 1  
AUTHORS Kaneko,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiumi,T.,  
Sasamoto,S., Watanabe,A., Idesawa,K., Iriguchi,M., Kawashima,K.,  
Kohara,M., Matsumoto,M., Shimpo,S., Tsuruoka,H., Wada,T., Yamada,M.  
and Tabata,S.  
Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
Bradyrhizobium japonicum USDA110  
DNA Res. 9 (6), 189-197 (2002)  
MEDLINE 22484998  
PUBMED 12597275  
REFERENCE 2  
AUTHORS Kaneko,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiumi,T.,  
Sasamoto,S., Watanabe,A., Idesawa,K., Iriguchi,M., Kawashima,K.,  
Kohara,M., Matsumoto,M., Shimpo,S., Tsuruoka,H., Wada,T., Yamada,M.  
and Tabata,S.  
Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
Bradyrhizobium japonicum USDA110 (supplement)  
DNA Res. 9 (6), 225-256 (2002)  
MEDLINE 22485002  
PUBMED 12597279  
REFERENCE 3 (bases 1 to 300600)  
AUTHORS Kaneko,T.  
Direct Submission  
TITLE Submitted (20-NOV-2002) Takakazu Kaneko. Kazusa DNA Research  
Institute, The First Laboratory for Plant Gene Research; 2-6-7  
Kazusa-kamatari, Kisarazu, Chiba 292-0812, Japan  
(E-mail: kaneko@kazusa.or.jp,  
URL: http://www.kazusa.or.jp/rhizobase//,  
Tel: 81-438-52-3935 (ex.2338), Fax: 81-438-52-3934)

FEATURES  
sourceLocation/Qualifiers  
1. 300600

/organism="Bradyrhizobium japonicum USDA 110"

/mol\_type="genomic DNA"

/strain="USDA110"

/db\_xref="taxon:224911"

/complement(27..917)

/gene="bll14605"

/complement(27..917)

/gene="bll14605"

/note="ORF\_ID:bll14605"

hypothetical protein"

/codon\_start=1

/transl\_table=11

/protein\_id="BAC49870.1"

/db\_xref="GI:27352873"

/translation="MFDECTMSKTTTIAILAPGANGSAVARRLSENGARVLTSLK

GRSSTAKRAADAGMGADDEAIAEADIILISIVPPGEAVALAEALALIVRRAKKPVV

VDCNAVNDTVLRIEEIGSAQAPFDVAGIIFPQPGKSPAFYMSGEALRGADVAVLK

ELGDMRIVEGPVGAALKMSYAGIVKGLAGIGSAMVIAATKAGAADALDEALSQ

PAVLARLEVALPDMIPKAYRWVAEMREISGFLGADHPASQIYEGFARWFELAEADAG

EAVDAGLWKAFASIAQKKA"

complement(1002..1772)

/gene="bll14606"

/complement(1002..1772)

/gene="bll14606"

/note="ORF\_ID:bll14606"

hypothetical protein"

/codon\_start=1

/transl\_table=11

/protein\_id="BAC49871.1"

/db\_xref="GI:27352870"

/translation="WTFRDDAPALDPIPEHTGLMGRIRNYELTGLVVTGPIAITLY

LVWFEVWDSQVRRFPFLAYRPPTYLPYVPGGLIVAFPTLILVGLFLANLIGRL

VDVGETFUGRIIPAVRIYGLKQVFETLPGKSSFRKVGLEFPSPGMWSIVLISQ

PNEVARSPOQEEHVSVFLPCSPNPTTGFYFPKSKIVEVDNMSTEDAAATLMSAGV

VQGAADPKKAALAGANAARIANASALQOPAKVE"

complement(1852..3678)

/gene="glms"

/complement(1852..3678)

/gene="glms"

/note="ORF\_ID:bll14607"

/codon\_start=1

/transl\_table=11

/product="glucosamine-fructose-6-phosphate

aminotransferase"

/protein\_id="BAC49872.1"

/db\_xref="GI:27352871"

/translation="MCGIVGILGREPVAEQVLDSILKRLRYGYDSAGVATLEGRHLER

RRAEGLKNLEKLEAEPLKGTGTGHTRWATHGKPTVNNAPHATERVAVWHNGIIE

NFRLELEKNGTVFHTETDTEIVLHLDLLTRGNKPVFAVKLALARLGAALGP

IFAGDNDLMICARNGPPLAIGYGGENYLGSDAIALGFTDTISYLEDDGWVLTNRN

ATIFDKGHAOREKIRHAASTSLVDKANYRHFHFAKEIHEQPEVVGHTLARYVDYATE

RVSLFVLFPDFKPIQIRNITACGTASYAGFAVKYFWRFAFLPEVDVDAVEFVREA

PLRGDLAIFISQSGETADTLAALRYAKAGVHTVAVNVVPTSTIARESVTLPTLAG

PEIGVASTKATCTOLMLANLAIAGARGELSDDETKLVHGLVEIPLMSDALTTE

LQIEKLAREIAKSDVLYLGRGTSPPLALEGALKLSIYIHAEGYAAGELKHGPIAL

LIDETPVVVIAPYRVEFKTVSNQEVNARGGKILLMTDAKGAEEATVESLTVIWPFD

MAAATPMVAVPVQVOLLAYHTAVTGTDDVDQFRNLAKSVTVE"

complement(4121..5476)

/gene="glmu"

/complement(4121..5476)

/gene="glmu"

/note="ORF\_ID:bll14608"

/codon\_start=1

/transl\_table=11

/product="UDP-N-acetylglucosamine pyrophosphorylase"

/protein\_id="BAC49873.1"

/db\_xref="GI:27352872"

/translation="MTAKRSSLTIIVLAGEGTRMRSHLPKVLHPVAHOTLLAHVLTAAAP

KGTGTSVAVTGPQVAADEARRIRPDVTFVQAERLGTAHVLAALAEATARGVDDL

LIAFGDTPLISAEFTARLAPLANGAALAFRAADPAGYGRFIVEGDRILVAIREQA

DASADERKIDLCNAGVMAIDGRRALAILDKIGNANSKEYVLTDAVEIVRQGWESVY

gene

CDS

5649..7028

/gene="bll14609"

5649..7028

/gene="bll14609"

/note="ORF\_ID:bll14609"

hypothetical protein"

/codon\_start=1

/transl\_table=11

/protein\_id="BAC49874.1"

/db\_xref="GI:27352873"

/translation="MAKDSPLADAGTGTGTGGLFSGLLAEESAFDRMMWRLLSGWGV

AAGVAVVYMNQAQGLWRDDQNASADLVQRADRLQLLTRESONEARRLAAAVETLN

ADRLRYSVTVLEQGLDVSVTGAIRQSAKQPDTPAPQAPAAPSVAPASMPSPSSD

KPRAAAKEQVKEATKDTVKETAKESPPPSAAASLVQOSPATTALPMLPLVPS

KSIWAPDEPAASKLTQPTETKETETAEKKPEPTAPTEITAAAKPPPEATESEAPA

IAYQOTRFATDLGANSIDGLRALMRGVTKSNPEVATLRPIIMIKEGTTGLGMOLRLG

AGLLINAAAALCAGLAENDRHCEITVFDQRLSMRGSEKGEKQSDRVQEKNAEK

NPDAAPQAEATAPAPAPTAKPKRRRSYSSKRSSKREEPAPAPAPAPAKPETAAAG

STLSSPFR"

complement(7106..7345)

/gene="bsl14610"

complement(7106..7345)

/gene="bsl14610"

/note="ORF\_ID:bsl14610"

hypothetical protein"

/codon\_start=1

/transl\_table=11

/protein\_id="BAC49875.1"

/db\_xref="GI:27352874"

/translation="MADGAEQMDVEQVQDRKSFVEFVAAMRNELLVGAKGWNVDLAD

FLSAMTAADHSHSPANPNPWRHAAALMWAGKLYE"

7466..7846

/gene="bll14611"

7466..7846

/gene="bll14611"

/note="ORF\_ID:bl14611"

hypothetical protein"

/codon\_start=1

/transl\_table=11

/protein\_id="BAC49876.1"

/db\_xref="GI:27352875"

/translation="MTKKPRLTPYQVQLIVGVFVAVGYAIYLRYLGIENSNVGLAC

DAGLQTLMKCARSVSTVLFKNSVFGIVALVATLNLMRPSIVLLTVGVVAAGLGIVLY

NVVLGSLAIGLLILGFARPAPATA"

complement(7806..9470)

/gene="ndvC"

complement(7806..9470)

/gene="ndvC"

/note="ORF\_ID:bll14612"

putative beta (1-6) glucans synthase"

/codon\_start=1

/transl\_table=11

/protein\_id="BAC49877.1"

/db\_xref="GI:27352876"

/translation="MAREQVIDGASRRTEPISLRTPLALLLVSIGATAAVMMWLATPI

TLARADIDPADKQCVSYAPFGEQSPNANWTHIEADQIEODLRQKEITDCVRYTSM

ENGLDOVPAIAAKVGGKVLQGLWLSNNRAKNFEQVALAVRLTKEYPDIIITSVTVGNE

VLLRGENTADLVSFIRAVKQVVRVPYADVNEFWLRNREVTEAVDFVTHILPYME

DVPVRAKAAHVEQIRERMAVAPPGKEIILGETGWPSEGRMGALPSSRTNQARVVS

EILSLAKAQKFRVNLIESYDQPKWRKLEGTAGGYWGLYDSVRRNLKYVPGPEISNFPY

WKYMGAMGMLCALVFAVAITLRRRPWTPRFSTWGLVAISATTAGVLLGVGADKVVY

EYSGWGLHWMGLAAGILSPIFCAQALVIGRSLPTFLDLGLGPEGRKWSKLTAVLG

LTVAVTVIARQATALLAFVDFPRYHDFPYAALTMAVVPFALMLNRPQIGVYRPIAEAVF

ACVLASAAVYVLFNEGRDNWQSLWTCMYLLFALTMLRAREQSQE"

9694..10290

/gene="bll14613"

9694..10290

/gene="bll14613"

```

/note="ORF_ID:bl4613
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC49878.1"
/db_xref="GI:27352877"
/translation="MRORVRSRNVAVLRQFAATLAVSLIALAGLAGASAQSGTPAPD
OGKAAAPADAAADKAAQNRRTDEFAEAQVINGPAGNPECVWLGRRVLRVLRWDRDL
DTAFRLDLVDRFCPGGHIQAAFRCLTRFGGQIDPKVATLDSRHACWINPASQPO
QAAAAQAPAAPTSGNAQAQPPAASPPAPQK"
gene
10509..13178
/genes="ndvB"
CDS
10509..13178
/genes="ndvB"
/note="ORF_ID:bl4614"

```

## Alignment Scores:

```

Pred. No.: 1,85e-57 Length: 300600
Score: 766.00 Matches: 189
Percent Similarity: 53.40% Conservative: 125
Best Local Similarity: 32.14% Mismatches: 201
Query Match: 27.57% Indels: 73
DB: 1 Gaps: 12

```

US-09-868-987-14 (1-552) x AP005952 (1-300600)

```

Qy 1 MetValSerProfileLeuAsnValProLeuLys---AsnHisAlaSerValSerGly 19
Db 143619 GTGATCTCGCGCTGTCTCGCGAGCGGATCACCAGCGGTACAGGCGCGATCTCCGC 143560
Qy 20 LysPheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSer 39
Db 143559 AATTTCACCGTGAATCGGCCAACGATCTCGGATCTCTCGCGCGCGCGCTGCCG 143500
Qy 40 PheValProGluValLeuSerGluLysThrLysSerSerAspLeuGlyLysLysGlnCys 59
Db 143499 GCGCGCTCACCGGTGTCAGGAGCGACCGCTCGGTCCGGCGCTTGGCCAGACTCGATC 143440
Qy 60 ThrGlnGlyIleLeuSerAlaCysCysGlyLeuAlaMetLeuLeuValLeuMetSerVal 79
Db 143439 GAGAGGCGAGCTTGGCTTACGTGCGTTCGATCTCGATCATCTTCATCTCGTCTG 143380
Qy 80 TyrTyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuLeu 99
Db 143379 ACCTATCGGTGTTCGGGTGTTCGCCAACATCGCGTGGCCATCAACGTCGCCATGATC 143320
Qy 100 TrpAlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleVal 119
Db 143319 TTCGGCTGTCTGCTGCTCAACGCCACCGTCAAGCTGCGCGCATCGCGCATCGTG 143260
Qy 120 LeuAlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGlu 139
Db 143259 CTCACGTCGGCATCGCGGTGGACTCAACGTCGTGATCTACAGCGCATCCCGGAGGAG 143200
Qy 140 PheLeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGly 159
Db 143199 CTGCGCGCGCGCGTAAACGATCTCGCGATCGACCGCGCTTCAAGCGCGCTCGCG 143140
Qy 160 AlaIlePheAspSerLeuThrValLeuAlaSerAlaLeuLeuPhePheLeuAsp 179
Db 143139 ACCATCTCGATTCCAAATCAACACCTTCATTGCGCGCGGTGTGTCTACATCGGC 143080
Qy 180 ThrGlyProIleLysGlyPheAlaLeuThrLeuLeuLeuGlyIlePheSerSerMetPhe 199
Db 143079 ACCGCGCGGTGCGCGCTTCGCGGTGACGCTCGGTATCGGCATCATCACCACGGTTC 143020
Qy 200 ThrAlaLeuPheMetThrLysPhePhePheMetLeuTrpMetAsn----- 214
Db 143019 ACCGCTTCAACCTGACCGCGGTGATCTGTCGCGCGGTGGTGGTGAAGCGCGCGCAG 142960
Qy 215 -----LysThrGln-HisThrGln-HisThrGln-HisMetMetAsnLysPh 227
Db 142959 AGCGTCCGATCTAGGAGCGCTGATGTGGAATCACACCGTTCTCATCGGCTCGCGCTCCT 142900

```

```

Qy 227 eValGlyIle----- 230
Db 142899 CATTACTATTCTCACCGTGTTCGCGGCCCTCGCGCTTGTGCTGCCGTCTGTCGCATCGTCCC 142840
Qy 231 -----LysHisAspPheLeuArgGlyCysLysLysLeuTrpAlaValSerGlySe 247
Db 142839 GGACAACACGCAATTCGACTTCACGCGATTCCGCGCATCAGCTTCCCGATCTCCGCGC 142780
Qy 247 rValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerValLeuGlyMe 267
Db 142779 GCTGTGATCGTCCGATCACGCTGTTTACCCACGCGCTGAAC-----CTCGGTAT 142726
Qy 267 tAspPheLysGlyGlyTyrAlaPheThrPheAsnProLysGluHisGlyIleSerAspVa 287
Db 142725 CGACTTCAAGGCGCGCAGCTTGTGAGGTGCGGCGCAGTCC---GGCGCCCGCAGCAT 142669
Qy 287 lAlaGlnMetArgGlyLysValValHisLysLeuGlnGluAlaGlyLeuSerSerArgAs 307
Db 142668 CGCGCGCATGCGCACGCGCTG-----GATGGCTTGGCTCGGCGCA 142627
Qy 307 pPheArgIleGlnThrPheGlySerSerGluLysIleLysIleTyrPheSerAspLysAl 327
Db 142626 GTCCAGCTCCAGCAGTTCGCGCGCGCGAGAACGTCCTGATCCGCTCGCGGAGCAGCC 142567
Qy 327 aLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeuThrIleMetSerTrpAr 347
Db 142566 GCGCGCGGATACCGCCACGAGGCGCGTG----- 142536
Qy 347 gTyrCysGlyIleValValArgAsnArgProArgPheLeuTyrGlyAsnSer----- 364
Db 142535 -----CAGAAAGTCCGTACCGCGCTCGCGGATTCATCCAGTA 142498
Qy 365 -LysArgAsnAlaLysPheTrpSerLysValSerLysLeuSerLysLysMetArgTy 384
Db 142497 CCGCGCGCTCGAGGTCTCGTCCGCGCTCGCGCGCAGCTC-----TTGCGCTA 142447
Qy 384 rGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeuLeuValSerLeuAr 404
Db 142446 C---GGCATGCTCGGCTGATCTCGGATCTCGGATCTGATCTATCTCTGTGTTCCG 142390
Qy 404 gPheGluTrpGlnTyrAlaPheSerAlaValCysAlaIleuIleHisAspLeuAlaTh 424
Db 142389 CTTCAATGGCAGTTCGCGCTCGCGCATGATCGCAGCGTGCAGCAGATC----- 142338
Qy 424 rCysAlaValLeuPheIleAlaHisPhePheLeuLysLysIleGlnIleAspLeuGlnAl 444
Db 142337 -----GTGCTGACGATCGGCTTCATGTCGATCAGTCAGTTCGATTCGACCTGACCAG 142285
Qy 444 aIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIlePheAs 464
Db 142284 CATCGCGCGCTTCGACCATCTCTGGCTATTTCGCTTACAGCAGCGTCTGATCTACCA 142225
Qy 464 pArgIleArgGluAspArgGlnAlaAsnLeuPheThrProMetHisValLeuValAsnAs 484
Db 142224 CGGTATCGTGAATGCTGCGCGCTACAGAAGATGCGATGCGCGAGCTGCTCAACCA 142165
Qy 484 pAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThrThrLeuSerVal 504
Db 142164 GTCCATCAACTCGACGCTGTCGCGTTCGATCATCACCGACGTCAGCTGCGCTG 142105
Qy 504 uLeuMetLeuLeuPheIleGlyLysSerValPheAsnPheAlaPheIleMetThrI 524
Db 142104 GTCGCGCTGCTGCTGTCGCGGTCAGCCATCCACAGCTTCACCGCGGTGATGATGTT 142045
Qy 524 eGly---lLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProProLeuLeuLeuPh 543
Db 142044 CGCGCTGTGTGTCGTCGCGACCTACACCTCGATCTTCATCGCGCGCGCATCTGATCTA 141985
Qy 543 eMetValArgLysGluAsnArg 550
Db 141984 TCTCGCGCTCGCGGAGCATCGC 141963

```

```

RESULT 12
AE004799/c
LOCUS
DEFINITION
Pseudomonas aeruginosa PA01, section 360 of 529 of the complete
genome
ACCESSION
AE004799 AE004091
VERSION
AE004799.1 GI:9949981
SOURCE
Pseudomonas aeruginosa PA01
ORGANISM
Pseudomonas aeruginosa PA01
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas
1 (bases 1 to 14652)
Stover,C.K., Phan,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K.-S., Wu,Z. and Paulsen,I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
Nature 406 (6799), 959-964 (2000)
20437337
10984043
2 (bases 1 to 14652)
Stover,C.K., Phan,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
Hancock,R.E.W., Lory,S. and Olson,M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics
University of Washington Genome Center, University Of Washington,
Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 14652)
Pseudomonas aeruginosa Community Annotation Project (PseudocAP)
Direct Submission
Submitted (04-FEB-2003) Department of Molecular Biology and
Biochemistry, Simon Fraser University, 8888 University Dr.,
Burnaby, British Columbia V5A 1S6, Canada
-----
This represents the February 3, 2003 version of the continually
updated, reviewed, Pseudomonas aeruginosa PA01 genome annotation,
from PseudocAP (see http://www.pseudomonas.com for latest updates
and links to alternate annotations). PseudocAP is coordinated by
Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert
E.W. Hancock (University of British Columbia, Canada). We welcome
submission through www.pseudomonas.com of any proposed changes.
'Protein name confidence' is used to rate our confidence of the
accuracy of the protein name.
Class 1: Function experimentally demonstrated in P. aeruginosa.
Class 2: Function of highly similar gene experimentally
demonstrated in another organism (and gene context consistent
in terms of pathways its involved in, if known).
Class 3: Function proposed based on presence of conserved amino
acid motif, structural feature or limited sequence similarity
to an experimentally studied gene.
Class 4: Homologs of previously reported genes of unknown function,
or no similarity to any previously reported sequences.
-----
Location/Qualifiers
1. .14652
/organism="Pseudomonas aeruginosa PA01"
/mol_type="genomic DNA"
/strain="PA01"
/db_xref="taxon:208964"
complement(153..353)
/locus_tag="PA3808"
/note="synonym: yfhj"

FEATURES
source
gene
complement(153..353)
/locus_tag="PA3808"
/note="Protein name confidence: Class 4 (homologs of
previously reported genes of unknown function, or no
similarity to any previously reported sequences)"
/transl_table=1
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG07195.1"
/db_xref="GI:9949982"
/translation="MSLKWTDLVLEIAIQLDSEKPEVDPRVNFVDLRRWVVELPEFSD
DPQSGEKVLEIAQAAWIEERD"
complement(380..718)
/gene="fdx2"
/locus_tag="PA3809"
complement(380..718)
/gene="fdx2"
/locus_tag="PA3809"
/note="Protein name confidence: Class 2 (high similarity
to functionally studied protein)"
/codon_start=1
/transl_table=1
/product="ferredoxin [2Fe-2S]"
/protein_id="AAG07196.1"
/db_xref="GI:9949983"
/translation="MPQIVILPHADHCHPEGAVFEAKPGKETILDAAALRNGIEIEHACEK
SCACTTCHVIVREGLDSMEPSDELEDDMLDKAWGLEPSRLSCQAVVADEDLVVEIPK
YTINOVSEGH"
complement(725..2584)
/gene="hscA"
/locus_tag="PA3810"
complement(725..2584)
/gene="hscA"
/locus_tag="PA3810"
/note="Protein name confidence: Class 2 (high similarity
to functionally studied protein)"
/codon_start=1
/transl_table=1
/product="heat shock protein HscA"
/protein_id="AAG07197.1"
/db_xref="GI:9949984"
/translation="MALLQIAEPQSPKPHERRLAVGIDGTTNSLVAAVRSGVAEPL
PDAGRLILPSAVRYHAERAEVGSERAAAAADPFNTVISVKRLMGRGLGVKOLGEO
LPYFRGEGSHMPFIETVOGLKSEVSAADILRELRORAETGLLGLGVAVITVPAYF
DDAQRATKDAARLGLNVLRLNEPTAAAVAYGLDKAGELVAIYDLGGCTFDISIL
RLTGVEFVLTGSDTALGGDDFHAHAGVIEQAGLSADLDPSGQQLLIACAKE
RLTDEASRVAYGWSGELSRATLIEFFVAHSKSCRRVRDSDGVDLDEIRKSVV
VGSTRVPRVRTAVGELFCEPLTDIDPDQVVAIGAIAIQADLAGNKRGEELLDDVI
PLSLGLTMGLMEKVIPTNTTIPVRAQFTTYKDGQTAMMIHVLOQERELVKDCRS
LARPELGIIPMVAGAAKIRVTFQVDADGLLGVSAELSSGEASIQVKPSYGLTDGE
IARMLKDSFDYAGDDKAARLREQVQAORLLEAVQSALDVGDERLDEERLAAQ
MDTLRELAGSDTAAIENQIKRLSQVTDFAARMMDATVKAALSGRRLNEIEE"
complement(2627..3148)
/gene="hscB"
/locus_tag="PA3811"
complement(2627..3148)
/gene="hscB"
/locus_tag="PA3811"
/note="Protein name confidence: Class 2 (high similarity
to functionally studied protein)"
/codon_start=1
/transl_table=11
/product="heat shock protein HscB"
/protein_id="AAG07198.1"
/db_xref="GI:9949985"
/translation="MGKPCFHAQFDLQPAFLVDLDELQGRYRELVRSHVDFRFADEP
REORLALERAQNLNEAYTQKSAPRRALYLLTSGHELPEATVQDREFLLQOQLRE
ELELODSADLAGVATPKRLKAAQAELEREFAACMDDAQORREAEERLVRMQFLDKL
AQEVRLQEERLDD"
complement(3156..3479)
/gene="iscA"
/locus_tag="PA3812"
complement(3156..3479)

```

/gene="isca"  
 /locus tag="PA3812"  
 /note="Protein name confidence: Class 3 (function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene)"  
 /codon\_start=1  
 /transl\_table=1  
 /product="probable iron-binding protein IscA"  
 /protein\_id="AAG07199.1"  
 /db\_xref="GI:9949986"  
 /translation="MWSLSTAAAKHVSLEGRKGEGIRLGRVITTCGSLAYVLFVDELAEDLVFESHGVKVIIDPKSLVLDGTDELFTREGLNEGFNNPNVRGCGCGESFNV"

## gene

complement (3507..3893)  
 /gene="iscU"  
 /locus tag="PA3813"  
 /gene="iscU"  
 /locus tag="PA3813"  
 /note="Protein name confidence: Class 3 (function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene)"  
 /codon\_start=1  
 /transl\_table=1  
 /product="probable iron-binding protein IscU"  
 /protein\_id="AAG07200.1"  
 /db\_xref="GI:9949987"  
 /translation="MAYSEKVIDHYENPRNVGKLDADPNVGTOMVGAPACGDVWRILQIKVNEQVIEDAKFTYCGGSAIASSLATEMMKGLTDEATIKNTTIAEELALPPV KIHCSVLAEADAKAAVDYRKOKGLL"  
 /complement (3929..5143)  
 /gene="iscS"  
 /locus tag="PA3814"  
 /complement (3929..5143)  
 /gene="iscS"  
 /locus tag="PA3814"

## CDS

complement (3507..3893)  
 /gene="iscU"  
 /locus tag="PA3813"  
 /gene="iscU"  
 /locus tag="PA3813"  
 /note="Protein name confidence: Class 3 (function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene)"  
 /codon\_start=1  
 /transl\_table=1  
 /product="probable iron-binding protein IscU"  
 /protein\_id="AAG07200.1"  
 /db\_xref="GI:9949987"  
 /translation="MAYSEKVIDHYENPRNVGKLDADPNVGTOMVGAPACGDVWRILQIKVNEQVIEDAKFTYCGGSAIASSLATEMMKGLTDEATIKNTTIAEELALPPV KIHCSVLAEADAKAAVDYRKOKGLL"  
 /complement (3929..5143)  
 /gene="iscS"  
 /locus tag="PA3814"  
 /complement (3929..5143)  
 /gene="iscS"  
 /locus tag="PA3814"

## gene

complement (3929..5143)  
 /gene="iscS"  
 /locus tag="PA3814"  
 /note="Protein name confidence: Class 2 (high similarity to functionally studied protein); Alt protein name: iscS protein"  
 /codon\_start=1  
 /transl\_table=1  
 /product="L-cysteine desulfurase (pyridoxal phosphate-dependent)"  
 /protein\_id="AAG07201.1"  
 /db\_xref="GI:9949988"  
 /translation="MKLPIYLDYSATTVPDPRVAQKMSCELLMDGNFGNPARSHVFGWKAEAVENARQVAADPREIVMTSGATESDNIAIKGVAHFYSGKGKHIITSKI EHKAVLDTCQLREGPEVTVLEPEDGLITPALVEAALRDDTILVSMVHWNIEITV NDIAAIGELTRSRVLPHVDAQSTGKVEIDLKLDVLSFSAHKTYGPKGICALYV RRPVRLEGOMHGSGHERGSRGTLATHQIVGMGEAFRIAKSEMAQENARVIALDR PFPAQIDGLEBYINGMTSRVPHNLNVSFNVEGESLIMALKOLAVSSGSACTASLDE PSYVRLALGRNDELHAHSIRFTGFRFTTEBEIDYAAKKVVEAVSKURELSPLMDMYKE GVDLSQVWQAH"

## CDS

complement (5173..5664)  
 /gene="iscS"  
 /locus tag="PA3814"  
 /note="Protein name confidence: Class 2 (high similarity to functionally studied protein); Alt protein name: iscS protein"  
 /codon\_start=1  
 /transl\_table=1  
 /product="L-cysteine desulfurase (pyridoxal phosphate-dependent)"  
 /protein\_id="AAG07201.1"  
 /db\_xref="GI:9949988"  
 /translation="MKLPIYLDYSATTVPDPRVAQKMSCELLMDGNFGNPARSHVFGWKAEAVENARQVAADPREIVMTSGATESDNIAIKGVAHFYSGKGKHIITSKI EHKAVLDTCQLREGPEVTVLEPEDGLITPALVEAALRDDTILVSMVHWNIEITV NDIAAIGELTRSRVLPHVDAQSTGKVEIDLKLDVLSFSAHKTYGPKGICALYV RRPVRLEGOMHGSGHERGSRGTLATHQIVGMGEAFRIAKSEMAQENARVIALDR PFPAQIDGLEBYINGMTSRVPHNLNVSFNVEGESLIMALKOLAVSSGSACTASLDE PSYVRLALGRNDELHAHSIRFTGFRFTTEBEIDYAAKKVVEAVSKURELSPLMDMYKE GVDLSQVWQAH"

## gene

## Alignment Scores:

Pred. No.: 7,17e-59 Length: 14652  
 Score: 762.50 Matches: 182  
 Percent Similarity: 55.42% Conservative: 130  
 Best Local Similarity: 32.33% Mismatches: 183  
 Query Match: 27.45% Indels: 68  
 DB: 1 Gaps: 13

US-09-868-987-14 (1-552) x AE004799 (1-14652)

Qy 1 MetValSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys 20  
 Db 10764 ATCATCAGCTGGCGACCATCCAGTCGCGCTGGGCAACAGTTCCGTATCACCAGGCTC 10705  
 Qy 21 PheThrHisArgGluValSerLysLeuAlaSerAspLysSerGlyAlaMetSerPhe 40

Db 10704 GATGGCCCGGCGAGTCTCCGAGCTGGCCCTGCTGCTGCTGCGCGCGCTGGCTGCG 10645  
 Qy 41 ValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysLysGlnCysThr 60  
 Db 10644 CCGATGTACTTCCGCGAAGAACACCATCGCGCCGAGCTGGGCGGCGAACAACATCGCC 10585  
 Qy 61 GlnGlyIleIleSerAlaCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80  
 Db 10584 AAGGTATCGATCGCTCGCTCGGCGCATGCTCTTGTATCGCTGTTCATCATCGTCTATC 10525  
 Qy 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuLeuLeuTyr 100  
 Db 10524 TACCGTTTCTTCGCGGTATCGCCACCGCTTGCCTTCAACATGGTTCATGCTGCTG 10465  
 Qy 101 AlaAlaLeuGlnTyrIleuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120  
 Db 10464 GCCTGATGTCGATCTCGCGCGCATCCCTGACCTCGCGCATCGCGGTATCGTGTCTG 10405  
 Qy 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleAtrGluGluPhe 140  
 Db 10404 ACCATGGCATGGCGGTGGAGCCACAGTCTGATCTTCTCGGCGATACCGGAGGAACTG 10345  
 Qy 141 LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160  
 Db 10344 GCCAATGGCATGTCCGTCCAGAGGGGATCCACGAAGGCTTCAACGCTGCTTCCACGCG 10285  
 Qy 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180  
 Db 10284 ATCTCGACGCCAACCTGACCTCGCTGCTGGCGGATCTCTACGCCATCGGCGACC 10225  
 Qy 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200  
 Db 10224 GGCCCGGTGAAGGATTTCCTGCGGACCATGCTGCTCGGATCATCACCTCGATGTTTCA 10165  
 Qy 201 AlaLeuPheMetThrLysPhe-----PhePheMet 210  
 Db 10164 GCCATCATGTTCACCGCGCAATGGTCAACCTGATCTTCGGCGGCGTGACTTCAAGAA 10105  
 Qy 211 LeuTyrMet-AsnLysThrGln---HisThrGlnLeuHisMetMetAsnLysPheValG 229  
 Db 10104 CTGTGATCTTAAGGGGACTGAGATGAATATCAAGATCGGTACCATCAAC---TTCAATGG 10048  
 Qy 229 YlleLysHisAspPheLeuArgGlyCysLysLysLeuTyrAlaValSerGlySerValPh 249  
 Db 10047 CATCCGCAAC-----GTGCGCTTTCGCGTCC 10021  
 Qy 249 eLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTyrPheSerVal-----LeuG 266  
 Db 10020 CTTGATC---CTGACCGTATCGCCCTGGCGAGCTGTTTCAACAGGGCATCAACTTCGG 9964  
 Qy 266 YMetAspPheLysGlyGlyTyrAlaPheThrPheAen?oLysGluHisGlyLysSerAs 286  
 Db 9963 CTTGACACTTACCGCGGCTACGCTGATCGAGTGTACCTACGAACAACCG-----GCCGA 9910  
 Qy 286 pValAlaGlnMetArgGlyLysValValHisLysLeuGlnGluAlaGlyLeuSerSerAr 306  
 Db 9909 TCTCGCAAGTTCGCTGGTCACTGCTC-----GGCGCCGCTATGA 9868  
 Qy 306 gAspPheArgIleGlnThrPheGlySerSerGluLysIleLysIleTyrPheSerAspLy 326  
 Db 9867 GGACGCGGTGTCAGAGCTTCGGCGACGCGCGAGCTGCTGCTGCGGATGCCAGCGCA 9808  
 Qy 326 sAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeuThrIleMetSerTr 346  
 Db 9807 GGACCGGAACCTGGGCAAGAAGTTCCTACCGCTTGCAGCGCGCAT-----9759  
 Qy 346 pArgTyrCysGlyIleValValArgAsnArgProPheLeuTyrGlyAsnSerLysAr 366  
 Db 9758 -----GCCGCGCAACCGCGC 9745  
 Qy 366 gAsnAlaLysPheThrSerLysValSerSerLysLeuSerLysLysMetArgTyrGlnAl 386  
 Db 9744 GAACCTGAAGCGGTTCAGTACGTACGCTCGCGCGCGAGGTGGTGAAGAACTCGCGCAGCGG 9685

Qy 386 aThrileGlyLeuLeuGlyAlaLeuAlaIleileLeuLeuTyfValSerLeuAtrpHeG1 406  
 Db 9684 CGGCTCGGCATGCTCTCGGCGTGGCGGCATCTCTGTACGTGCGCTCCGCTTCCA 9625  
 Qy 406 uTrpGlnTyAlaPheSerAlaValCylAlaLeuIleHisAspLeuLeuAlaThrCysAl 426  
 Db 9624 GTGGAAGTTCCGCTCGGTCGATCTCTCGTGTGACGATGCGATCGTATGGG 9565  
 Qy 426 aValLeuPheIleAlaHisPhePheLeuLysIleGlnIleAspLeuGlnAlaIleG1 446  
 Db 9564 CTGCTG-----TCGTTCTTC-----CAGGTGACCTTCGACCTGACCGTCTCTCGC 9520  
 Qy 446 yAlaLeuMetThrValLeuGlyTyfSerLeuAsnAsnThrLeuIleIlePheAspArgI1 466  
 Db 9519 CGCTGCTGGCGGTGGTGGTACTCTGCTGAACGACACCATCGTGATCTTCGACCGGT 9460  
 Qy 466 eArgGluAsp-----ArgGlnAlaAsnLeuPheThrProMetHisValLeuVa 482  
 Db 9459 CGCGGAGAACTTCCGCGTCTCGCAAGGCCGATCTGGTTCGAGAACCTGAACATC----- 9405  
 Qy 482 IaAsnAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThrThrLeuSe 502  
 Db 9404 -----TCCACACCGACACCTGTTGCTACCATCGCCACCTCGGTTTCCACGCTCT 9352  
 Qy 502 rValLeuLeuMetLeuPheIleGlyGlySerSerValPheAsnPheAlaPheIleMe 522  
 Db 9351 GCGTATCGCGCGCTGCTGTTCTTCGGCGGCGACATCTGTTGCTGTTCTCCATCGCCT 9292  
 Qy 522 tThrileGlyIleLeuLeuGlyThrLeuSerSerLeuTyfIleAlaProProLeuLeuLe 542  
 Db 9291 GTTCGCTGTAATGCGCGGTACCTACTCGTCGATCTACATCGCAACGCTGTTATTCAT 9232  
 Qy 542 upheNet 544  
 Db 9231 CTGGCTG 9225

RESULT 13  
 AE016777  
 LOCUS Pseudomonas putida KT2440 section 4 of 21 of the complete genome.  
 DEFINITION  
 ACCESSION AE016777 AE015451  
 VERSION AE016777.1 GI:26557020  
 KEYWORDS  
 SOURCE Pseudomonas putida KT2440  
 ORGANISM Pseudomonas putida KT2440  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 1 (bases 1 to 300861)  
 Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,  
 Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,  
 Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,  
 Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,  
 Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Uterback,T.,  
 Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,  
 Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,  
 Dueterhoft,A., Tummeler,B. and Fraser,C.  
 Complete genome sequence and comparative analysis of the  
 metabolically versatile Pseudomonas putida KT2440  
 Environ. Microbiol. 4 (12), 799-808 (2002)  
 2 (bases 1 to 300861)  
 Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,  
 Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,  
 Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,  
 Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,  
 Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Uterback,T.,  
 Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,  
 Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,  
 Dueterhoft,A., Tummeler,B. and Fraser,C.  
 Direct Submission  
 Submitted (05-NOV-2002) The Institute for Genomic Research, 9712  
 Medical Center Dr, Rockville, MD 20850, USA  
 Location/Qualifiers

source

1. 300861  
 /organism="Pseudomonas putida KT2440"  
 /mol\_type="genomic DNA"  
 /strain="KT2440"  
 /db\_xref="taxon:160488"  
 114. 2351  
 /locus\_tag="pp0788"  
 114. 2351  
 /locus\_tag="pp0788"  
 /note="identified by Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAN66413.1"  
 /db\_xref="GI:24982192"  
 /translation="MATSIWRHSFDDFSALSCKTPCKESRMOKEGKVVPPLAATIDRGG  
 RLPHCLPVLQVRDKAQLQKQLODLFDNADDTLUFEMADKAFDGGDQNLVPEAMR  
 DLRKRKSIERGEFLDTFDAPLAGQVLDLANLAQPGKPRSKAQMERAALIEGVAVR  
 LSRDGVALLQGLRLQALLDRPLYEYHNPGLCPAALCRYFLDAGNLGVRLKVLVK  
 LFERYVLDVDVITYGEANQLAAAGVLPQLQAPRRRAEDRRMRVRPGPANLGOEBA  
 DAGOALFASLQALLAPVRCGFAPRLQAAQAQPISTADLLRLSLHLQRYVPATYEA  
 DFELGQQLLELLSVRSVRSRRRLHVAEDMNLVGLLFAFIHNDNLPLASRALIT  
 RLHPLIKVALLDKGLFSRSSHPARRLLNEIAGAAIGWEGSECCORSLMRVERIVQ  
 RLNDFADDAASLFAELDDLELAFQDERRERNELEQRTDAEGRARMAQAOQVHV  
 LNORLRGRVLPQVVQMLVOAQSVQLLAWLKQGEASQAWRDAVQTMMLIASITPPH  
 EPOALQOVPLGLKALRDGLASVALDAVATREFFLQELHLRACACKEALPSGEGS  
 LQVNLVQEELVLTIAEPPACVPLHTADQAAALRQAOQLRIGTHVHEVLDEDEPLRCKL  
 VARIDSSDLRVFANRTGMKRVWSSASLAQALQHEVRVLDRLFLFRALEAVLESLR  
 GQGAQ"  
 2440. 3012  
 /gene="ampD"  
 /locus\_tag="pp0789"  
 2440. 3012  
 /gene="ampD"  
 /locus\_tag="pp0789"  
 /note="similar to GP:4079828, GB:K02765, SP:P01024,  
 PID:I79665, and PID:I81130; identified by sequence  
 similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="N-acetyl-anhydromuramyl-L-alanine amidase AmpD"  
 /protein\_id="AAN66414.1"  
 /db\_xref="GI:24982194"  
 /translation="MQLDYATGWFHIGITHCPSNPNARPEGESISLLVTHNISLPP  
 ACSTGKVKQOFFQNRRLAPDEHPYFASINHLTVSAHLVERDGAQTQVSLHDRAWHAG  
 VSFDRGREGNDFSIGIEGTDLDPYDAQYAVLEQLTHRSAPFAIDUNRTIQGHS  
 DIAPQRKTDGPAFDWSRYRQALQHNEDKA"  
 3009. 3839  
 /locus\_tag="pp0790"  
 3009. 3839  
 /locus\_tag="pp0790"  
 /codon\_start=1  
 /transl\_table=11  
 /product="inner membrane protein AmpE"  
 /protein\_id="AAN66415.1"  
 /db\_xref="GI:24982195"  
 /translation="MSFLVLLALWKEFSALRHQVORDGFFIGELVRLERNKGVHPW  
 WTALLVAPVALLVLLHVEPVAYGLLAPVHLLVLIYSLGRGDANASLUGAPRDW  
 RRGDDQAAHAAERDLGVADPHSLVRYQGNLLMQVYQGFYFVYVFGPAAL  
 AYRLALCSEHSKQPALKARAEQRLIMDLVPVRLALSALVGNFLAVTRVLMHLEVL  
 NWHISAGHLVARVGRITADTIPEEDSQRLGLRLDSLWELLRLCAVLWVAGFALWTVLV  
 "  
 complement(3884. 4660)  
 /locus\_tag="pp0791"  
 complement(3884. 4660)  
 /locus\_tag="pp0791"  
 /note="identified by match to TIGR protein family HMM  
 TIGR00010"  
 /codon\_start=1  
 /transl\_table=11  
 /product="deoxyribonuclease, TatD family"  
 /protein\_id="AAN66416.1"  
 /db\_xref="GI:24982196"

gene

CDS

gene

CDS

gene

CDS

gene

CDS

/translation="WRLDTHLDFPDDADRLANAAAGLERMVLGVYQANF  
QVWELACAEPLFAALGLHPVYLDQHRPEHLVOLREWLRLNPRLCVAGRGGLDY  
YLALDKTRQALFEAQMACDFELPALLVRRSHQVYIATLKRYPARAGVYIHPA  
GSEYBEARYIKGLFLLGGAGTWPQALRLKRTIARLPESVLEITSDPMAPAMYGP  
VRNPEHLPEIATALAHVMGUDVEVLAQASTRNACTLFGW"  
complement(4663.5658)

gene

/gene="frub"  
/locus\_tag="PP0792"  
/complement(4663.5658)  
/gene="frub"  
/locus\_tag="PP0792"  
/note="similar to GP.9949716; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="fructose transport system repressor Frub"  
/protein\_id="AAN66417.1"  
/db\_xref="GI:24982197"

CDS

/translation="MKLSDIARLAGSVTTASVYVINGKAEQORISNSTVRRVAVFA  
HGFTPNQAGLSRHRRTILGFIPLDENPSYARIKQEGGARAGYQLLIASSDDQ  
PDSERQQLFRARCDALFVASCPEPDDSYRELQDKGUPVIAIDRRLPAPFCSVI  
SDDRAGROLAASLSAPRSIALIGARPELSVSQARAGGFDEALQGYTGEVRYQGE  
AFSDEQRQLMQIIDLGLPDLVTTSYVLLQGVFTLQARVPDSRQLQGLTFGN  
QLLDFLPVNNAMAQQHQAATALELALAAIEBKRYEPGVHAGVTRFKQRIISA"  
5943.8795  
/gene="frub"  
/locus\_tag="PP0793"  
5943.8795  
/gene="frub"  
/locus\_tag="PP0793"  
/note="similar to SP.P23388; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="phosphotransferase system, fructose-specific  
EI/HPr/EIIA components"  
/protein\_id="AAN66418.1"  
/db\_xref="GI:24982198"

gene

/translation="MLELANQIANGKQKADKABALLLADRLVADGLVAGYLOGLQ  
AREAGSTFLGQGAIPHGTPTQRLDYATGRLQFPFEGVMDGQGVYLAIGIAR  
SDEHLRLQLTALGTDIAELARRASSAEALLKLQGAPOALALDAQLVGLNPAE  
DFDLWRGLARLQRCADVDSFAALVQQAEPPLGBGLMHLHSEROVROPGLAFITP  
QOPLRYRDPNLGFCILAGAAHEALLERLCEVLEIRGOVLYQATSSRAVLEVLGG  
EVPADPSPARVLANPGLHARPAKLAQLAKFEGEIRVRLVDSQAAPVKSLSKL  
LSICARGQALELVAESFIADALPVLIAETQGLGEEVEPLQSVAPTDVPEVLQ  
APAGSRIGVGAAPATGAPHVCFEREFDYPLRGESCALERKUREALATVNGELQ  
ALVLRSDKATGEIIVTHQEMLADPALTDDVEQLAQGESAAAMWAVIEAARQEQAL  
HDALLABRAADRLRIGRVLQALCCVQAQVEPQYVYLVMTVEPVSARLDPNRVAG  
IVTAQGGATAHSATVARALGIPAVVGAGASTILLESCTPLLDGQGVSVAGAEVGL  
ORALERDLRQRLQAAWAFPEALTDDGHAVEVFANIGDSNGIAKVPQGAEGVL  
LRTLEIEMAHPOAPDVATQAEYRRVLDGLDRLPVLRTLDVGDKPLPYWPIAAEEN  
PFLGVGRVRLTLQPQVMEQDLRALRAADQRPRIIMPVQGVHWEAREARVERLR  
AELIVADQLQIMVEPVSAAALAAQAREVDFFSITGNDLTQYTLAIDRGHPSLSQA  
DGLHPVLSLIDMTVRAAHAGKHWGVGCLAADPQAVAVLLGLDVLDELVSARSATAE  
VKALVROADHQTARALAREALQDSAAAVALVERY"  
8796.9743  
/gene="frub"  
/locus\_tag="PP0794"  
8796.9743  
/gene="frub"  
/locus\_tag="PP0794"  
/note="similar to GB.L20852, and PID:306772; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="1-phosphofructokinase"  
/protein\_id="AAN66419.1"  
/db\_xref="GI:24982199"

CDS

/translation="MAKILTLTNLPALDITIGLDTLRPGQVNSRQASHAAKGLNV  
AQVLDLGHVSVVGFGGDNLPQFEALIDGRGFTDCFVRVPGETSNIKLVRADGRV  
TDINGQGVDEARSALLRLHQEALFAGHDAAVVVAGSLPRGISADFRQLRLKQAG  
LKVYLDGSGALRYGLQSPWLKPNTVEELGVLGAVDNLTOQRAAAKRLDSGVGH  
VVSAGSGQVSWFSDRLAQARPPKRVASVGVGDSLVAGMVHGLLLAEAPQTLTR

gene

/translation="WRLDTHLDFPDDADRLANAAAGLERMVLGVYQANF  
QVWELACAEPLFAALGLHPVYLDQHRPEHLVOLREWLRLNPRLCVAGRGGLDY  
YLALDKTRQALFEAQMACDFELPALLVRRSHQVYIATLKRYPARAGVYIHPA  
GSEYBEARYIKGLFLLGGAGTWPQALRLKRTIARLPESVLEITSDPMAPAMYGP  
VRNPEHLPEIATALAHVMGUDVEVLAQASTRNACTLFGW"  
complement(4663.5658)

CDS

ATAIAQAQVTVQVGFHIDREHLAQLAEAAVQLTEQEGCR"

gene

9740.11482  
/locus\_tag="frua"  
/locus\_tag="PP0795"  
9740.11482  
/gene="frua"  
/locus\_tag="PP0795"  
/note="similar to GB.J04213, GB.L34219, SP.P12271,  
PID:L90854, and PID:S98229; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="phosphotransferase system, fructose-specific  
IIBC component"  
/protein\_id="AAN66420.1"  
/db\_xref="GI:24982200"

CDS

/translation="WNIAIVTACPNQGVSVLSARLLSAAQRGMSTSEVQNTFHP  
ERQLSAAQIAEADWLVVSTGPDLARFVGLKYQSTPSQALADRGFLDEAANAEL  
LAAVTCAPAAASSAGARI VAVTACPTGVATFMAAEALQAAQQLGYLTVETQGSVG  
ARNPLSAEALAAADVLLAADIEVTPARFAGKRIYRCGTGIALKQARATLDKALAQT  
4.92e-55 Length: 300861  
Pred. No.: 739.50 Matches: 184  
Score: 54.53% Conservative: 123  
Percent Similarity: 32.68% Mismatches: 184  
Best Local Similarity: 26.62% Indels: 72  
Query Match: 1 Gaps: 13  
DB: 1

Alignment Scores:

US-09-868-987-14 (1-552) x AE016777 (1-300861)

|    |       |  |       |
|----|-------|--|-------|
| Qy | 1     | MetValSerSerProIleuAsnValProLeuLysAAsnHisAlaSerValSerGlyLys  | 20    |
| Db | 69350 | ATCATCAGCTGGCGACCATCCAGTCGCGCTTGGTAGGACAGTTCGCGATCACCGCTTG   | 69409 |
| Qy | 21    | PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe | 40    |
| Db | 69410 | AATGGCCAGGGCAATCGTCCGAGCTGGCCCTGCTGCTGCGTCCGCTGGTCTGGCTGCA   | 69469 |
| Qy | 41    | ValProGluValLeuSerGluGluThrIleSerSerAspLeuLysLysGlnCysThr    | 60    |
| Db | 69470 | CCGATGTACTTCGCTCAGAACAGTACCATTTGGCCCAAC3CTGGGTCTGCTGACATCACC | 69529 |
| Qy | 61    | GlnGlyIleSerAlaCysGlyLeuAlaMetLeuLeuValLeuMetSerValTyr       | 80    |
| Db | 69530 | AAGGGTATCGATCGCTGCTGGGCGATGCTTTCGCTCGCTGTTTCATCATCGCCATC     | 69589 |
| Qy | 81    | TyrArgPheGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuLeuTyr       | 100   |
| Db | 69590 | TACCGCGCTTCGGTGTGTGGCCACCATCGCCCTGGCGGCAACATGGTGTGCTGCTG     | 69649 |
| Qy | 101   | AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu | 120   |
| Db | 69650 | GCCTTCATGCTGCTGCTGGGGCGACCTGACCTGCGCGGTATTGCGCGCATCGTGTG     | 69709 |
| Qy | 121   | AlaMetGlyMetAlaValAspAlaValLeuValPheGluArgIleArgGluGluPhe    | 140   |
| Db | 69710 | ACATCGGTATGGCGGTGGACCCACGTCGTGATCTTCGCGTATTTCGGAAGAGATC      | 69769 |
| Qy | 141   | LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla    | 160   |
| Db | 69770 | GCGGTGTTGATGTGCTGGTGGCGGCATCCACGAGGTTCAATCGCCCTACTCTCGGCA    | 69829 |
| Qy | 161   | IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr | 180   |
| Db | 69830 | ATCGTCGACGCCAACCTGACCACTGTTGGTGGCGCGCATCTGTTTTCGCGGTAC       | 69889 |
| Qy | 181   | GlyProIleLysGlyPheAlaLeuThrLeuLeuGlyIlePheSerSerMetPheThr    | 200   |
| Db | 69890 | GGCCCCGTCAAAGGCTTTCGGTCCACATGTCCTCGCGGATTTTACCTCGATGTTTACC   | 69949 |
| Qy | 201   | AlaLeuPheMetThrLysPhePhePhe-----Met                          | 210   |



[illegible]

|            |   |                    |                         |        |                 |
|------------|---|--------------------|-------------------------|--------|-----------------|
| RESULT 14  | AE012971/c  | 7687 bp            | DNA                     | linear | BCT 01-JUL-2002 |
| LOCUS      | Chlorobium tepidum TLS  | section 192 of 194 | of the complete genome. |        |                 |
| DEFINITION | AE012971 AE006470   |                    |                         |        |                 |
| ACCESSION  | AE012971.1  | GI:21648307        |                         |        |                 |
| VERSION    |   |                    |                         |        |                 |
| KEYWORDS   |   |                    |                         |        |                 |
| SOURCE     | Chlorobium tepidum TLS  |                    |                         |        |                 |
| ORGANISM   | Chlorobium tepidum TLS  |                    |                         |        |                 |
| REFERENCE  | Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae; Chlorobium.   |                    |                         |        |                 |
| AUTHORS    | 1 (bases 1 to 7687)   |                    |                         |        |                 |
|            | Eisen, J.A., Nelson, K.E., Paulsen, I.T., Heidelberg, J.F., Wu, M., Dodson, R.J., Deboy, R.J., Gwinn, M.L., Nelson, W.C., Haft, D.H., Hickey, E.K., Peterson, J.D., Durkin, A.S., Kolonay, J.L., Yang, F., Holt, I., Umayam, L.A., Mason, T., Brenner, M., Shea, T.P., Parksey, D., Nierman, W.C., Feldblyum, T.V., Hansen, C.L., Craven, M.B., Radune, D., Vamathevan, J., Khouri, H., White, O., Gruber, T.M., Ketchum, K.A., Venter, J.C., Tettelin, H., Bryant, D.A. and Fraser, C.M. |                    |                         |        |                 |
| TITLE      | The complete genome sequence of Chlorobium tepidum TLS, a photosynthetic, anaerobic, green-sulfur bacterium   |                    |                         |        |                 |
| JOURNAL    | Proc. Natl. Acad. Sci. U.S.A. 99 (14), 9509-9514 (2002)   |                    |                         |        |                 |
| PUBLISHED  | 12093901  |                    |                         |        |                 |
| REFERENCE  | 2 (bases 1 to 7687)   |                    |                         |        |                 |
| AUTHORS    | Eisen, J.A., Nelson, K.E., Paulsen, I.T., Heidelberg, J.F., Wu, M., Dodson, R.J., Deboy, R.J., Gwinn, M.L., Nelson, W.C., Haft, D.H., Hickey, E.K., Peterson, J.D., Durkin, A.S., Kolonay, J.L., Yang, F., Holt, I., Umayam, L.A., Mason, T., Brenner, M., Shea, T.P., Parksey, D., Nierman, W.C., Feldblyum, T.V., Hansen, C.L., Craven, M.B., Radune, D., Vamathevan, J., Khouri, H., White, O., Gruber, T.M., Ketchum, K.A., Venter, J.C., Tettelin, H., Bryant, D.A. and Fraser, C.M. |                    |                         |        |                 |
| TITLE      | Direct Submission   |                    |                         |        |                 |
| JOURNAL    | Submitted (30-APR-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  |                    |                         |        |                 |
| REMARK     | unpublished   |                    |                         |        |                 |
| FEATURES   | Location/Qualifiers   |                    |                         |        |                 |
| source     | 1..7687   |                    |                         |        |                 |
|            | /organism="Chlorobium tepidum TLS"  |                    |                         |        |                 |
|            | /mol_type="genomic DNA"   |                    |                         |        |                 |
|            | /strain="TLS"   |                    |                         |        |                 |
|            | /db_xref="taxon:194439"   |                    |                         |        |                 |
|            | 163..780  |                    |                         |        |                 |
| gene       | /gene="rnhB"  |                    |                         |        |                 |
|            | /note="synonym: CT2261"   |                    |                         |        |                 |
|            | 163..780  |                    |                         |        |                 |
| CDS        | /gene="rnhB"  |                    |                         |        |                 |
|            | /notes="identified by match to PFAM protein family HMM PF01351"   |                    |                         |        |                 |
|            | /codon_start=1  |                    |                         |        |                 |
|            | /transl_table=11  |                    |                         |        |                 |
|            | /product="ribonuclease H11"   |                    |                         |        |                 |
|            | /protein_id="AA073476.1"  |                    |                         |        |                 |
|            | /db_xref="GI:21648308"  |                    |                         |        |                 |
|            | /translations="MLSTDFEYLPWESLSQVCGIDBAGRGLAGVVAAVVPRHFRPTGIFAKLDDSKLTAEELDELALRESAESWALGVDAETIDRINILQATMLANNLAVESLSTPEFLVLDGNRRPRVLPPIPIQYIVKDSKVFSAASVLAKTRHDELMTTS  |                    |                         |        |                 |
|            | AA073476.1  |                    |                         |        |                 |
|            | 800..1189   |                    |                         |        |                 |
| gene       | /gene="CT2262"  |                    |                         |        |                 |
|            | 800..1189   |                    |                         |        |                 |
| CDS        | /gene="CT2262"  |                    |                         |        |                 |
|            | /notes="identified by Glimmer2; putative"   |                    |                         |        |                 |
|            | /codon_start=1  |                    |                         |        |                 |
|            | /transl_table=11  |                    |                         |        |                 |
|            | /product="conserved hypothetical protein"   |                    |                         |        |                 |
|            | /protein_id="AA073477.1"  |                    |                         |        |                 |
|            | /db_xref="GI:21648309"  |                    |                         |        |                 |
|            | /translations="MAAPQLGAGEGKIAARHLAAAGYRIVARNYRFRHNEIDIIAFDGEALCTIEVKTRASKLGGHPAESVTSRQKEIAARAAAGYLAASLDDPWITCRFDVIAVIALSIDRSRTKSYEIEHIIKAFAMVGDGK"  |                    |                         |        |                 |
|            | 1262..3205  |                    |                         |        |                 |
| gene       | /gene="gvrB"  |                    |                         |        |                 |



```

Db      5225 GTTCCAAATGCGGCGTTCGCGGCATTTTTCACGACGTCGTGACGCGTCTCGGCGCTGT 5166
Qy      428 ---LeuPheileAlaHisPhePheLeuLysLysileGlnileAspLeuGlnAlaileG1 446
Db      5165 CAGCCTGCTGCGGCGCATTTTGACTTCATGCCGCTGGAGGTGCATCAGACGATATATCGC 5106
Qy      446 yAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuileilePheAspArgil 466
Db      5105 CGGCTTCTGACCATCGCGGTACTTCGATCACCACACAGTTCGTCGTATGACCGTAT 5046
Qy      466 eArgGluAspArgGlnAlaAsnLeuPheThrProMetHisValLeuValAsnAspAlaLe 486
Db      5045 CGCGAGCGGTGCGCGCGCCAGACCCCTCGGAGTACGAGCGGCTTTTCAACGAGAGCAT 4986
Qy      486 uGlnLysThrPheSerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuWe 506
Db      4985 GAACGACAGCGTTTCGCGCACCATCATCACCTCCGGCACCATCATGCTTCGTCGATCGT 4926
Qy      506 tLeuLeuPheileGlyGlySerSerValPheAsnPheAlaPheileMetThrileGlyil 526
Db      4925 GCTCTTCATCTTCGCGCGCGCGCCATCCGACGCTTCGCTTCGCCATGTTCCGCGGCAT 4866
Qy      526 eLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProProLeuLeu 541
Db      4865 TCTGTCGCGACCTACTCTCGATTTCTGTCGCGCACCGATTGTG 4820

RESULT 15
LOCUS   AE016860
DEFINITION Pseudomonas syringae pv. tomato str. DC3000 section 5 of 21 of the complete genome.
ACCESSION AE016860
VERSION   AE016860.1
KEYWORDS AE016860.1 GI:28851610
SOURCE  Pseudomonas syringae pv. tomato str. DC3000
ORGANISM Pseudomonas syringae pv. tomato str. DC3000
REFERENCE 1 (bases 1 to 311249)
AUTHORS Buell,R., Joardar,V., Khouri,H., Fedorova,N., Tran,B., Russell,D., Berry,K., Utterback,T., Van Aken,S., Feldblyum,T., Gwinn,M., Dodson,R., DeBoy,R., Durkin,A., Kolonay,J., Madupu,R., Daugherty,S., Brinkac,L., Beanan,M., Haft,D., Selengut,J., Nelson,W., Davidse,T., White,O., Fraser,C. and Collmer,A.
TITLE    Complete Sequence of Pseudomonas syringae
JOURNAL  Unpublished
REFERENCE 2 (bases 1 to 311249)
AUTHORS Buell,R., Joardar,V., Khouri,H., Fedorova,N., Tran,B., Russell,D., Berry,K., Utterback,T., Van Aken,S., Feldblyum,T., Gwinn,M., Dodson,R., DeBoy,R., Durkin,A., Kolonay,J., Madupu,R., Daugherty,S., Brinkac,L., Beanan,M., Haft,D., Selengut,J., Nelson,W., Davidse,T., White,O., Fraser,C. and Collmer,A.
TITLE    Direct Submission
JOURNAL  Submitted (03-MAR-2003) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES
Source  1..311249
         /organism="Pseudomonas syringae pv. tomato str. DC3000"
         /mol_type="genomic DNA"
         /strains="DC3000"
         /db_xref="taxon:223283"
         /note="pathovar: tomato"
         162..818
         /locus_tag="PSPT01158"
         162..818
         /locus_tag="PSPT01158"
         /locus_tag="PSPT01158"
         /note="identified by match to PFAM protein family HMM PF00440"
         /codon_start=1
         /transl_table=11
         /product="transcriptional regulator, TetR family"
         /protein_id="AAO54687.1"

gene
CDS

```

```

/db_xref="GI:28851611"
/translators="MVKTTREPSATSLKRRMRLMGKRTLLDAALEIFSRVYGHGSS
LDQVASLADVSKTNLYVPSKDDLVNLVRLLEVMWLSPLVHFTADKEVQATGAVI
KAKLEMRDHPAESRLFCMEVMOGAPLIQGELOHPLRDTVOAKAVVQIHWLDSOLAP
INPHLLIFTUWATTQHYADPRTQVEAVTGKTLDDPVFEFVSLRSRWLVDGLPRTA
"
1052..2143
/locus_tag="PSPT01159"
1052..2143
/locus_tag="PSPT01159"
/notes="similar to GP:13937517; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="bmp family protein"
/protein_id="AAO54688.1"
/db_xref="GI:28851612"
/translators="MYKKLLPGTLCAAVFAAGAAQAEGLTLKGPAKIAMVYISPRN
DGMWQAFDEARLKEKDLTKIQFVESPENAAATPVVDRLIARGANVIVGTAFGV
SDTFLAKKYPDVAFNGSGTTNAANLESFYGRITYESOYLCAAGAAKVTGKLGfV
AANPFGQVNTINAYELGAROTNPKATVTVTYTGAWNDPVKERAATWALIDNGVDVVG
OHVDSPTPOVAQERGIHGTGHRDLSEFAPKATVCCSSVMVMDRFLGPELKKIIAGNW
TPRANGALLSMQGGGTDTLTGDPFIISAENKKIEDERABELLAGKKIYISGPLEADRG
KERVAAQQUSDPDLWKMDWFVEGVKTQQ"
2140..3651
/locus_tag="PSPT01160"
2140..3651
/locus_tag="PSPT01160"
/notes="similar to GP:15023582; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding protein"
/protein_id="AAO54689.1"
/db_xref="GI:28851613"
/translators="MSQVNALQGTISKSPDFKALIDADFTACWGEVHALLGNGAG
KSLIMNIAGLYAPETGTLMLDDNPVLOQPKDASRYRIGMHQFKLVKVPFTVAQNI
LLALPEQDGSYRKLRLMLERDVRKAEELFAIDPSRVVDSLSVAEQORVELTKVL
LAGARILLDEPTAVLTDOEARLLITVQAFQGAIVILVTHKMAVDRKRVADRTVM
RGRTOTLPQRVSVEQLVLTGVSVPVAHHHPAVGVEVRLOKDLRSVGGGALNG
VNITLARGYIAGVGGNGQAEALNMGLPQATGEDIHLAPFGDLRNASAEORRQL
RNASIPADRYGAALAGLSVAENFGINSHYGSFWRIKGLKQDAQRAVESFDV
QGVRSIDQKALLSGGNAOKLVAAREFSDPQLVHSPSRGLDVRATQAVHDLRAA
RDAGAVLLISEDLDELVALADRIGVNMGGRIVAEFDHPADRAQIGKAMVSHD"
3644..3838
/locus_tag="PSPT01161"
3644..3838
/locus_tag="PSPT01161"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO54690.1"
/db_xref="GI:28851614"
/translators="MTLQAQPLTAPITARQPLLRRTLELRQGTGWPLOALJIALA
LVVGLAICTAILILAGVPAN"
3854..4764
/locus_tag="PSPT01162"
/notes="This region contains a gene with one or more
premature stops or frameshifts, and is not the result of a
sequencing artifact; ABC transporter, permease protein,
degenerate; similar to GP:15139998; identified by sequence
similarity; putative"
4784..5690
/locus_tag="PSPT01163"
/notes="This region contains an authentic point mutation,
causing a premature stop, and is not the result of a
sequencing artifact; ABC transporter, permease protein,
authentic point mutation"
5804..6493
/locus_tag="PSPT01164"
5804..6493
/locus_tag="PSPT01164"

```

```

/note="similar to GP:5123490; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ompA family protein"
/protein_id="AA054691.1"
/db_xref="GI:28851615"
/translation="MFTSRRLIIIVATAVALLSGCATSNPNYDNOGAOSSGOMSKTAKY
GGGLAGALAGAAIDHNNRCKGALIGAAGVAGASAGYGYADQAEALRASMANITGV
EVQGGQDIKILMFGNITFATDSALASSFYSPLNLSLKLQFNQNNIEIIGTDSY
GSRQNMDSLSQORASQSVATYLTSCQVDQSHLSVRGAGPDQPIASNADANGAQRNRVE
VNLKPIPGQQY"
6822..7922
/locus_tag="PSPT01165"
6822..7922
/locus_tag="PSPT01165"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA054692.1"
/db_xref="GI:28851616"
/translation="MFFHMLPWSWDSIRRLQDRDIAPYRGHFTPRNEVPVSPALIM
RFGNVSFAFDRAWEACRILGQRFVLDLSSIIKLDLIDVVTQAMTVVGSFAFVGGAA
GAGAGVLFVGAVGPGILMGAAALGAQASAFILGILGAEIAESVIDGPRRIIEYTRG
ITTAWQGPREGQRPFGWDDLAQNSAVOEIAQGHVWVILLGAMVOYITRGRGNAR
ALANEMASAGERLPGWLEHDELKRPDLORPEPRGALAEQETLSNRSKTADA
AKPDAMQKTVSSKIDPKQLDLVTGHNHSLSVHGVKSDGDLMLRAKSAIKPNQDV
GPSVPLSSAFHSDLELIFADQITRETLKKVS"
7963..8940
/locus_tag="PSPT01166"
7963..8940
/locus_tag="PSPT01166"
/note="similar to GP:13094159, and GP:4514540; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ISPssy, transposase"
/protein_id="AA054693.1"
/db_xref="GI:28851617"
/translation="MKQMTFADAEYAGKRRHRRERFLIEMQVVPWKGLIALIEPHY
PKGGSPAPYPLAMRIHLMQNWFGYSDPAMEEALYETTLIRQFAGLSLDRIDMETT
ILNFRLLKHELAGKQIOWINGSLVGRMLRGQITVDATIIHAPSSTKNKDKGRDP
EMHOTKNGNOYFFCMKAIQVDAESGLVSLVGTAAVADVTOVDOLHCEETVSGD
AGYTVGDKRAEHDROMWISIAAPPSRYKHGKSLIARVVRKIETTKAQLRAKVEHP
FRVTKRFGYTKVRFRGLAKNTAQQAATLFPALSNLMMWRKLLANGEVRL"
10294..10887
/locus_tag="PSPT01167"
10294..10887
/locus_tag="PSPT01167"
/note="similar to GP:15024601; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA054694.1"
/db_xref="GI:28851618"
/translation="METFGQVQEQVPESSIDRYRGKLPNQLLTYWAEHGWCGYGDG
IFVTNQEVDVAVVMELEGTKEFTFDAYHLIARSAFGDLYLWGEIEGASLKITSILS
RYTHKSLCIGSQMDKQLQSFILSRKVGSDNYGDLFRPARKLGLTHHDMYGFVPAL
MLGADTLDHLEKVKAVEHLVLAQLSDLESYDFQDL"
complement(11408..12331)
/locus_tag="PSPT01168"
complement(11408..12331)
/locus_tag="PSPT01168"
/note="identified by match to TIGR protein family HMM
TIGR00530"
/codon_start=1
/transl_table=11
/product="acyltransferase family protein"
/protein_id="AA054695.1"
/db_xref="GI:28851619"
/translation="MDFLPAPLRGVIALSLLALNTIVCCTVLPFAVTLKICLPFSAQ

```

```

RFTDMLSHIHETWIGTNNAWIRLCHTRWQLSGLEGLDYKHYSYLVTSNHSQWVDIMV
LOVVLNRRIRPLKFLKELIWPVIGLAWALQFPFMKRYSKAYLAKHPKKGKOLE
TTRTCTAKPRHNPVGIFPAEGTRFTAGKHQQQSPPRHLKPKAGGIAFVLDMGSEQ
LESIIINTIHPGGQFGYWDLLCGNVTDVVAHFEEIRIPQFLGKNYDQGEYRLFEQ
QWINALWEKDSLLDLRLEHQYPAQRGRDAF"
complement(12396..14486)
/gene="pta"
/locus_tag="PSPT01169"
complement(12396..14486)
/gene="pta"

```

## Alignment Scores:

| Pred. No.:             | 1.08e-54 | Length:       | 311249 |
|------------------------|----------|---------------|--------|
| Score:                 | 736.00   | Matches:      | 183    |
| Percent Similarity:    | 54.06%   | Conservative: | 123    |
| Best Local Similarity: | 32.33%   | Mismatches:   | 182    |
| Query Match:           | 26.49%   | Indels:       | 78     |
| DB:                    | 1        | Gaps:         | 16     |

US-09-868-987-14 (1-552) x AE016860 (1-311249)

| Qy | 1      | MetValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys  | 20     |
|----|--------|---|--------|
| Db | 284316 | ATCATCAGCTGCGCACCACCTCCAGTCCGCGCTGGCGAFTCAGTTCCGGATTACCGGTCTG | 284375 |
| Qy | 21     | PheThrHisArgGluValSerLysLeuAlaSerAspLysSerGlyAlaMetSerPhe     | 40     |
| Db | 284376 | AATGGTCAGGTGAGTCTCCGAGCTGGCGCTCTGTCGCGCGGTGGTCTGGCGCGG        | 284435 |
| Qy | 41     | ValProGluValLeuSerGluThrIleSerSerAspLeuGlyLysGlnCysThr        | 60     |
| Db | 284436 | CCGATGTACTTCGCTGAAGAGCGCACCATCGGTCCGAGCTCGGTCGCCACACATCACC    | 284495 |
| Qy | 61     | GlnGlyIleSerAlaCysGlyLeuAlaMetLeuValLeuMetSerValTyr           | 80     |
| Db | 284496 | AAGGATATCGACCTTCCTGTTGGGCGATGCTGCTGATCCATCTTCATCATGGCGATC     | 284555 |
| Qy | 81     | TyrArgPheGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTyr        | 100    |
| Db | 284556 | TACCGTCTTCGCTGATCGCTACCTGTCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTG     | 284615 |
| Qy | 101    | AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu  | 120    |
| Db | 284616 | GGCTCATGTCTTCTGCTGGTGGTCCACATGACCTCGCGGTATCCCGGTATCGGTG       | 284675 |
| Qy | 121    | AlaMetGlyMetAlaValAspAlaValLeuValPheGluArgIleArgGluPhe        | 140    |
| Db | 284676 | ACCATGGTATGGCGGTGGATGCCACGTGTTGATCTTCTCGCGGATCGCTGAAGATC      | 284735 |
| Qy | 141    | LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla     | 160    |
| Db | 284736 | GCCATGGATGACGGTGAACGGCTATCAACAGAGTTTCGATCGCTACACGGCG          | 284795 |
| Qy | 161    | IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPheLeuAspThr     | 180    |
| Db | 284796 | ATCCTCGACGCCAACCCTGACGAGCTGCTGGTCCGGCGGATCTCTTTCCGATGGCACC    | 284855 |
| Qy | 181    | GlyProIleLysGlyPheAlaLeuThrLeuLeuGlyIlePheSerSerMetPheThr     | 200    |
| Db | 284856 | GGCCCGCTCAAGAGGTTTCGCTGACCATGCTCGCTCGGATCTTTACCTCGATGTTTACG   | 284915 |
| Qy | 201    | AlaLeuPheMetThrLysPhePheMetLeuTyrMetAsnLysThrGlnHisThrGln     | 220    |
| Db | 284916 | GCCATCATGTCACCGCGC-----                                       | 284933 |
| Qy | 221    | LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys  | 240    |
| Db | 284934 | ---GCTATGTCACACCTGATCTACGGC---GGTCGTGACTTC-----AGAAG          | 284975 |
| Qy | 241    | LeuTyr-----AlaValSerGlySerValPheLeuLeuGly-----                | 252    |
| Db | 284976 | TTGTGGATTAAAGGGGCTGCCATGTGTACGTACCATCACTTTCATGGCGGTTCGCAATGTT | 285035 |

```
QY 253 -----CysValAlaLeuGlyPheGlyAlaTrp-----Asn 262
Db 285036 GCGTTCCGATCACCAGTCCTTACCGCAGCTGGCTTTCAGCTGTTCCATAAGGG 285095
QY 263 SerValLeuGlyMetAspPheLysGlyTyrAlaPheThrPheAsnProLysGluHis 282
Db 285096 CTCAACTTTGGTCTGGACTTACCGCGGCTACGCTCATCGAGCTGACCTACGAGCGTCT 285155
QY 283 GlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnAlaGly 302
Db 285156 -----GCCGATCTG-----GGCAAGTTCTGCAGGAGCTGTTTCGCCCGGC 285197
QY 303 LeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleTyr 322
Db 285198 TAC-----CACGAAGCTGTTGTCAGAGCTTTGGCGCACGACTGCTGCTGGTGGC 285251
QY 323 PheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeuThr 342
Db 285252 ATGCCGGGTGAAGACCCGCAATTGGGCACCTAGGTCGCTGAAGCGCTGCGCAAG----- 285305
QY 343 IleMetSerTrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeuTyrGly 362
Db 285306 -----GCCGTGCCGATACCCCGCAGTGGTC----- 285332
QY 363 AsnSerLysArgAsnAlaLysPheTrpSerLysValSerSerLysLeuSerLysLysMet 382
Db 285333 -----AAGCGTGTGAGTTC-----GTCCGCCCGCAGGTCGGCGAAGAGCTG 285374
QY 383 ArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleIleLeuTyrValSer 402
Db 285375 CGTGATCAGGGTGGCTCGCATGTGATGGCGCTGGCGGTGTTCTGATCTACCTGGCG 285434
QY 403 LeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeu 422
Db 285435 TTCCGCTTCCAGTGAAGTTTCTCTCGCGGATCATTTGCTGATCCACGACGTGGTC 285494
QY 423 AlaThrCysAlaValLeuPheIleAlaHisPheLeuLysLysIleGlnIleAspLeu 442
Db 285495 GTGACGATGGGTATCCTG-----TCGTTCTTC-----CAGATTACCTTCGACCTG 285539
QY 443 GlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIle 462
Db 285540 ACGGTATTGGCGCGGTGCTGGCGATCATCGGTACTCGCTGAACACACCATTTGTGTC 285599
QY 463 PheAspArgIleArgGluAsp-----ArgGlnAlaAsnLeuPheThrProMet 478
Db 285600 TTCGACCGGGTTCGGAACACTTCGCGCTGTCGCCAAGGCTTCGCTGATCGAGAAC--- 285656
QY 479 HisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrAla 498
Db 285657 -----ATCAACATCTCCACCACCGACAGACACTGTCGCCACCATTTGCCACCTCGGT 285707
QY 499 ThrThrLeuSerValLeuLeuMetLeuLeuPheIleGlyGlySerSerValPheAsnPhe 518
Db 285708 TCGACCTGTTGGCGATCTCTGCACTGTGGCGGTTTCGGTGGCGACAGCTGGAAGGCTC 285767
QY 519 AlaPheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaPro 538
Db 285768 TCCATTGGTTGTTTCATCGGTGTCGGGGTACCTTACTCGTGTATCATCGCCAAC 285827
QY 539 ProLeuLeuLeuPheMet 544
Db 285828 GTGGTGTGATCTGGCTG 285845
```

Search completed: October 28, 2003, 17:40:24  
Job time : 7083 secs

**THIS PAGE BLANK (USE70)**

```

FT XX                               /*tag= a
PN WO200039158-A1.
XX PD                               06-JUL-2000.
XX PF                               99WO-CA01230.
XX PR                               98US-0113280.
XX PR                               98US-0113281.
XX PR                               98US-0113282.
XX PR                               98US-0113283.
XX PR                               98US-0113284.
XX PR                               98US-0113285.
XX PR                               98US-0113285.
XX PR                               98US-0113385.
XX PR                               98US-0114050.
XX PR                               98US-0114056.
XX PR                               98US-0114057.
XX PR                               98US-0114058.
XX PR                               98US-0114059.
XX PR                               98US-0114061.
XX PA (CONN-) CONNAUGHT LAB LTD.
XX PI Murdin AD, Oomen RP, Wang J;
XX PI WPI; 2000-452369/39.
XX PR P-PSDB; AAY95543.
XX PT Novel Chlamydia polynucleotides and polypeptides useful for diagnosis,
XX PT prevention and treatment of Chlamydia infection in mammals .
XX PS Claim 2(a); Page 55-58; 215pp; English.
XX CC The present sequence is that of Chlamydia pneumoniae genomic DNA
XX CC including an open reading frame that codes for CPN100686 RY 54 (see
XX CC AAY95543), a putative 93 kDa outer membrane protein. It is an example
XX CC of C. pneumoniae polynucleotide molecules of the invention (see
XX CC AA50030-42) that encode antigenic polypeptides (see AAY95543-55) useful
XX CC in the diagnosis, treatment and prevention of Chlamydia infection.
XX CC The polynucleotides can be utilised: in the recombinant production
XX CC of Chlamydia antigens using transformed unicellular host cells; in
XX CC vaccines or live vaccine vectors; in naked form or formulated with
XX CC a delivery vehicle for therapy and prophylaxis of Chlamydia
XX CC infection; and as probes or primers for diagnosis of Chlamydia
XX CC infection.
XX SQ Sequence 1864 BP; 498 A; 346 C; 401 G; 619 T; 0 other;

Alignment Scores:
Pred. No.:      2 27e-301      Length:      1864
Score:          2778.00      Matches:      552
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:            21      Gaps:      0

US-09-868-987-14 (1-552) x AAAS0030 (1-1864)
Qy      1 MetValSerSerProileLeuAsnValProleuLysAsnHisAlaSerValSerGlyLys 20
Db      101 ATGTCAGCAGCCCTATTTTAAACGTCCTCCATGAAATCATGCCAGTGCTCAGGAAA 160
Qy      21 PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe 40
Db      161 TTTACCCACCGTGAAGTGAAGAGCAACTCGCCTCAGATTAAATCATGAGCGATGCTCTTT 220
Qy      41 ValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysGlnCysThr 60
Db      221 GTTCCCGAGGTTCTCAGTGAAGAGACGATCTCTCTGATCTTGGGAAACAAATGATACA 280
Qy      61 GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 80

```

```

Db      281 CAAGGCATTATCTCAGCATGCTGCTGGCTTGGCAATGCTTATTGTTTGGATGAGCGTATAT 340
Qy      81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTyr 100
Db      341 TATAGATTGGAGGCGTCATCGCTTCGGGAGCTGTTCTTCTGAATCTTTGCTTATCTGG 400
Qy      101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120
Db      401 GCAGCTCTACAGTATTGGATGGCCACTCACCCTGTGTCAGGACTCGCTGGGATTGTTCTT 460
Qy      121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPhe 140
Db      461 GCTATGGGATGGCCGTAGATGCAAAATGTTCTGTATTGCAAAAGATCCGAGAGGAATTT 520
Qy      141 LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160
Db      521 TTATTGTCCTCAAGTCTTAAATAAATCTGTAGAAAAGAGATATACCAGGCTTTTGGAGCC 580
Qy      161 IlePheAspSerAsnLeuThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180
Db      581 ATTTTGAATCTAACTTGACTACAGTATTGGCTCAGCAGCTCTTTCTTCTAGATACA 640
Qy      181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200
Db      641 GGGCTATTAAAGGGTTGCTTTGACATTGATTTAGGAATTTTCTCTTCAATGTTTACG 700
Qy      201 AlaLeuPheMetThrLysPhePheMetLeuTyrMetAsnLysThrGlnHisThrGln 220
Db      701 GCTCTTTTTCATGACTAAATTTTCTTCATGCTGTGGATGAATAAGACCCCAACATACACAG 760
Qy      221 LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys 240
Db      761 TTGCATATGATGATTAAGTTCGTGGGATTAAGCATGATTTCTTGAGAGGATGCAAAAAA 820
Qy      241 LeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla 260
Db      821 CTTTGGGCTGTTTCTGGAAGTGTCTTTCTTTAGGTGCGTCTCTCGGTTTGGAGCC 880
Qy      261 TrpAsnSerValLeuGlyMetAspPheLysGlyTyrAlaPheThrPheAsnProLys 280
Db      881 TGGAAATTCGCTTTTGGGAATGGAATTTAAAGGAGGATGCGCTTTTACCTTTTAAATCAAAA 940
Qy      281 GluHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGlu 300
Db      941 GAGCATGGCATCAGCCATGTTGCTCAATGCGTGGCAAGTGTGTCATAAATACAGGAA 1000
Qy      301 AlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysLysLys 320
Db      1001 GCTGGTCTTTCTCTAGAGACTTCCGTATTCAAAACATTTGGATCTTTCAGAAAAAGATCAAA 1060
Qy      321 IleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLys 340
Db      1061 ATCTATTATTAGTATAAAGCTTTTAAAGCTATCTAAGCAGATACGAGCTCTCTCCTAAAA 1120
Qy      341 LeuThrIleMetSerTrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeu 360
Db      1121 TTAACCATCATGAGCTGGCGTATTGCTGGGATTGTTGTCAAGAACAGGCTAGATTCTC 1180
Qy      361 TyrGlyAsnSerLysArgAsnAlaLysPheTrpSerLysValSerSerLysLeuSerLys 380
Db      1181 TACGGAACCTCTAAACGAAACGCAAAATTTTGGTCAAGGTAAGCAGCAAACTATCGAAG 1240
Qy      381 LysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyValAlaLeuAlaIleLeuLeuTyr 400
Db      1241 AAAATGCGTTATCAGCGACCATCGGCGCTTTTAGGAGCTTTGGCAATCATCTTGTCTCTAT 1300
Qy      401 ValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAsp 420
Db      1301 GTGAGTTTGGCTTTGAATGGCAATATGCTTTCAGTGGCGTATGCGCTTTAATTCATGAC 1360
Qy      421 LeuLeuAlaThrCysAlaValLeuPheIleAlaHisPhePheLeuLysLysLysLys 440
Db      1361 CTTTGGCTACCTGTGCGTCTTGTGTTATAGCACATTTCTTTTGAAGAAAAATTCAAATA 1420

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 28, 2003, 15:26:11 ; Search time 418 Seconds  
(without alignments)  
3564.808 Million cell updates/sec

Title: US-09-868-987-14

Perfect score: 2778

Sequence: 1 MVSSPILNVLKHNASVSGK.....SLYIAPLLLRMRKNSK 552

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\*\_p2n.model -DEV=xlh  
-Q=/Cyn2\_1/USPTO\_spool/US09868987/runat\_28102003\_152601\_23827/app\_query.fasta\_1.711  
-DB=N Geneseq\_19Jun03 -QMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09868987 -CGCN\_1\_0 @runat\_28102003\_152601\_23827 -NCPUP=6 -ICPU=3  
-NO\_WMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq\_19Jun03.\*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length  | DB ID | Description |
|------------|-------|---------------|---------|-------|-------------|
| 1          | 2778  | 100.0         | 1864    | 21    | AAA50030    |
| 2          | 2730  | 98.3          | 1330025 | 20    | AAX91990    |
| 3          | 1920  | 69.1          | 1038602 | 20    | AAX01425    |
| 4          | 719   | 25.9          | 1830121 | 17    | AAT42063    |
| 5          | 686.5 | 24.7          | 45613   | 22    | AAF28535    |
| 6          | 678.5 | 24.4          | 45613   | 22    | AAF28535    |
| 7          | 672   | 24.2          | 2943    | 20    | AAX91694    |
| 8          | 672   | 24.2          | 2955    | 20    | AAX91570    |
| 9          | 664   | 23.9          | 1584    | 24    | ABK74549    |
| 10         | 653   | 23.5          | 2292    | 25    | ABT15148    |
| 11         | 646.5 | 23.3          | 2226    | 22    | AAX52892    |
| 12         | 646.5 | 23.3          | 2301    | 24    | ABN91389    |
| 13         | 646.5 | 23.3          | 3511    | 22    | AAX54037    |
| 14         | 646.5 | 23.3          | 3705    | 22    | AAX54228    |
| 15         | 618.5 | 22.3          | 349980  | 21    | AAA81457    |
| 16         | 618.5 | 22.3          | 349980  | 21    | AAA81457    |
| 17         | 618.5 | 22.3          | 349980  | 21    | AAF21608    |
| 18         | 618.5 | 22.3          | 1437668 | 21    | AAA81490    |
| 19         | 618   | 22.2          | 495269  | 24    | ABG67195    |
| 20         | 618   | 22.2          | 3011208 | 24    | ABG69245    |
| 21         | 616   | 22.2          | 2944528 | 24    | ABA03041    |
| 22         | 612   | 22.0          | 5839    | 24    | ABQ70971    |
| 23         | 589   | 21.2          | 910715  | 20    | ABX20248    |
| 24         | 512   | 18.4          | 4403765 | 22    | AAI99683    |
| 25         | 509   | 18.3          | 80557   | 24    | ABX09142    |
| 26         | 509   | 18.3          | 4411529 | 22    | AAI99682    |
| 27         | 461.5 | 16.6          | 18477   | 23    | AAI99634    |
| 28         | 429   | 15.4          | 349980  | 22    | AAH68529    |
| 29         | 414   | 14.9          | 835     | 22    | AAF26302    |
| 30         | 408.5 | 14.7          | 1848    | 22    | AAF94404    |
| 31         | 408.5 | 14.7          | 1848    | 24    | ABK64968    |
| 32         | 393   | 14.1          | 1671    | 19    | AAI13967    |
| 33         | 386.5 | 13.9          | 1581    | 18    | ABQ25097    |
| 34         | 381   | 13.7          | 1848    | 24    | ABQ90076    |
| 35         | 368   | 13.2          | 1024    | 24    | ABX65642    |
| 36         | 368   | 13.2          | 1024    | 24    | ABX66280    |
| 37         | 368   | 13.2          | 1024    | 24    | ABX66285    |
| 38         | 365   | 13.1          | 1377    | 20    | AAI20685    |
| 39         | 358   | 12.9          | 2322    | 23    | AAS93101    |
| 40         | 357.5 | 12.9          | 2323    | 23    | AAS89009    |
| 41         | 357.5 | 12.9          | 2326    | 23    | AAS90384    |
| 42         | 350   | 12.6          | 1035    | 19    | AAX14487    |
| 43         | 346.5 | 12.5          | 1854    | 25    | ABZ38871    |
| 44         | 346.5 | 12.5          | 1857    | 24    | ABK54085    |
| 45         | 342   | 12.3          | 975     | 22    | AAF94409    |

ALIGNMENTS

RESULT 1  
AAA50030  
ID AAA50030 standard; DNA; 1864 BP.  
XX  
AC AAA50030;  
XX  
XX 10-OCT-2000 (first entry)  
XX  
XX DNA encoding Chlamydia pneumoniae antigen CPN100686 RY 54.  
DE  
DE CPN100686 RY 54; antigen; infection; diagnosis; therapy; vaccine;  
KW outer membrane protein; ds.  
XX  
XX Chlamydia pneumoniae.  
OS  
XX  
XX Key Location/Qualifiers  
FH 101..1759  
FT CDS



Qy 441 AspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeu 460  
 |||||  
 Db 1421 GATTGGCAAGCCATTGGTGCTTTAATGACTGATTGGGTATTCTATTAACAACACTTTG 1480  
 Qy 461 IleIlePheAspArgIleArgGluAspArgGlnAlaAsnLeuPheThrProMetHisVal 480  
 |||||  
 Db 1481 ATCATTTTGTATCGTATTTCGTGAAGATCGCAAGCGAACCTGTTTACCCCTATGCTGTT 1540  
 Qy 481 LeuValLeuAspAlaLeuGlnLysThrPheSerArgThrValMetThrAlaThrThr 500  
 |||||  
 Db 1541 TTAGTTAATGATGCCCTTCAAAGACGTTACGCCGACGGTAATGACAAACAGCTACAACT 1600  
 Qy 501 LeuSerValLeuLeuMetLeuLeuPheIleGlyGlySerSerValPheAsnPheAlaPhe 520  
 |||||  
 Db 1601 CTATCAGTTTGTGTAATGCTTTGTTATAGCGGCTCTCTGCTTTAATTTGCAATT 1660  
 Qy 521 IleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProLeu 540  
 |||||  
 Db 1661 ATTATGACCATAGGATTTCTTAGGAACCTTATCGTCTCTTTATATTGCACCACTCTG 1720  
 Qy 541 LeuLeuPheMetValArgGlyGluAsnArgSerLys 552  
 |||||  
 Db 1721 TTGTTGTTTATGGTCCGTAAGAAATCGCTCAAAA 1756

## RESULT 2

AAX91990/c  
 ID AAX91990 standard; DNA; 1230025 BP.

AC AAX91990;

XX 13-SEP-1999 (first entry)

DT Nucleotide sequence of the complete genome of Chlamydia pneumoniae.

DE Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope; ss.

XX Chlamydia pneumoniae.

XX W09927105-A2.

XX 03-JUN-1999.

XX 20-NOV-1998; 98WO-1B01890.

XX 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

XX (G8ST ) GENSET.

PA Griffiths R;

XX WPI; 1999-357842/30.

DR Genome sequence of Chlamydia pneumoniae

PT Claim 1; Page 291-611; 1912pp; English.

PS The present sequence represents the complete genome of Chlamydia  
 CC pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes  
 CC respiratory disease such as pneumonia and bronchitis and is thought  
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,  
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides  
 CC encoded by the open reading frames of the C. pneumoniae genome (see  
 CC AAY34584-Y35879) can be used in immunogenic compositions as vaccines.  
 CC Vectors containing C. pneumoniae nucleotide sequences can also be  
 CC used as immunogenic compositions, especially where the vector directs  
 CC the expression of a neutralising epitope of C. pneumoniae.

XX Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

## Alignment Scores:

|                        |           |               |         |
|------------------------|-----------|---------------|---------|
| Pred. No.:             | 1.38e-291 | Length:       | 1230025 |
| Score:                 | 2730.00   | Matches:      | 551     |
| Percent Similarity:    | 99.28%    | Conservative: | 0       |
| Best Local Similarity: | 99.28%    | Mismatches:   | 1       |
| Query Match:           | 98.27%    | Indels:       | 3       |
| DB:                    | 20        | Gaps:         | 0       |

US-09-868-987-14 (1-552) x AAX91990 (1-1230025)

|    |        |  |        |
|----|--------|--|--------|
| Qy | 1      | MetValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLys   | 20     |
| Db | 661906 | ATGTCAGCAGCCCTATTTTAAACGTCCTCCATGAAAATCATGCCAGTGTCTCAGGAAA     | 661847 |
| Qy | 21     | PheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe   | 40     |
| Db | 661846 | TTTACCACCCGCTGAAGTGAGCAAACTCGCTCAGATTTAAATCTGGAGCGATGCTTTT     | 661787 |
| Qy | 41     | ValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysGlnCysThr      | 60     |
| Db | 661786 | GTTCCTCCGAGGTTCTCAGTGAAGAGCGATCTCTTCATCTTGGGAAAAACAATGTACA     | 661727 |
| Qy | 61     | GlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr   | 80     |
| Db | 661726 | CAAGGCATTATCTCAGCATGCTGTGGCTTGCCAAATGCTTATTTGTTGATGAGCGTATAT   | 661667 |
| Qy | 81     | TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrp      | 100    |
| Db | 661666 | TATAGATTGAGGCGCTCATCGCTTCGGAGCTGTTCTTCTGAAATCTTTTTCCTATCTGG    | 661607 |
| Qy | 101    | AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu   | 120    |
| Db | 661606 | GCAGCTCTACAGTATTGCGATGCGCCACTCACCTTGTCCAGGACTCGCTGGGATGTTCTT   | 661547 |
| Qy | 121    | AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPhe   | 140    |
| Db | 661546 | GCTATGGGATGGCGGTAGATGCAAAATGTTCTTGTATTTCGAAAGAAATCCGAGAGGAATTT | 661487 |
| Qy | 141    | LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla      | 160    |
| Db | 661486 | TTATTGTCTCAAGTCTTAAAAAATCTGTAGAAAAAGATATACCAAGGCTTTTGGAGCC     | 661427 |
| Qy | 161    | IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr   | 180    |
| Db | 661426 | ATTTTTCATCTTAACCTTGACTACAGTATTGGCCTCAGCACTCTCTTTCTTCTTAGATACA  | 661367 |
| Qy | 181    | GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr   | 200    |
| Db | 661366 | GGGCCTATTAAAGGGTTTGCTTTGACATTGATTTTAGGAATTTTCTCTTCAATGTTTACG   | 661307 |
| Qy | 201    | AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisThrGln      | 220    |
| Db | 661306 | GCTCTTTTTCATGACTAAATTTTCTTCATGCTGTGGATGAATAAGACCCCAACATACACAG  | 661247 |
| Qy | 221    | LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys   | 240    |
| Db | 661246 | TTGCATATGATGAATAAGTTCGTGGGATTAAGCATGATTTCTTGAGAGGATGCAAAAAA    | 661187 |
| Qy | 241    | LeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla   | 260    |
| Db | 661186 | CTTTGGGCTGTTTCTGGAAGTGTTTTCTTTTAGTTGCGTGTGCTCTCGGTTTGGAGCC     | 661127 |
| Qy | 261    | TrpAsnSerValLeuGlyMetAspPheLysGlyTyrAlaPheThrPheAsnProLys      | 280    |
| Db | 661126 | TGGAAATTCCTGTTTGGGAATGGATTTTAAAGGAGGGTATGCTTTTACCTTTAATCCAAA   | 661067 |
| Qy | 281    | GluHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGlu   | 300    |
| Db | 661066 | GAGCATGCGATCAGCGATGTTGCTCAATCGTGGCAAAAGTTGTGTCATTAACACTACAGAA  | 661007 |
| Qy | 301    | AlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLys   | 320    |
| Db | 661006 | GCTGGTCTTCTTCTAGAGACTTCCGTATTCAACATTTTGGATCTTCTCAGAAAAAGATCAAA | 660947 |



|    |        |   |        |
|----|--------|---|--------|
| Qy | 181    | GlyProIleLysGlysPheAlaLeuThrLeuLeuLeuGlyIlePheSerSerMetPheThr         | 200    |
| Db | 992651 | GGACCTATCAAAAGGTTTCGCGCTCACTTTGATCATCGGGATTTCTCATCCATGTTCA            | 992710 |
| Qy | 201    | AlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisThrGln             | 220    |
| Db | 992711 | GCCCTGTTATGACGAATATTTCTTGTCAITTCGGTGCAGAAAACGAGAGAGACCCAA             | 992770 |
| Qy | 221    | LeuHisMetMetAsnLysPheValGlyIleLysHisAspPheLeuArgGlyCysLysLys          | 240    |
| Db | 992771 | TTGCATATGATGAATAAGTTTATGGGATCAAGCATAACTTCTTAAAAGAAATGTAACGT           | 992830 |
| Qy | 241    | LeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla          | 260    |
| Db | 992831 | TTGTGGGTAGTTTCTGGAGTAGTCTCTGTTTAGGTTGTGTAGGATTAAGGTTTCGTGCT           | 992890 |
| Qy | 261    | TrpAsnSerValLeuGlyMetAspPheLysGlyGlyTyrAlaPheThr                      | 276    |
| Db | 992891 | TGGGATTCGTTTTTGGGATGGATTTTAAAGGTGGATATGCATCTACITTAGATTCCGAT           | 992950 |
| Qy | 277    | -----PheAsnProLysGluHisGlyIleSerAspValAlaGlnMetArgGlyLys              | 293    |
| Db | 992951 | GTGTGTCAGTACAATCCAGAG-----CAGATGTGTTCTGTA                             | 992986 |
| Qy | 294    | ValValHisLysLeuGlnGluAlaGlyLeuSerSerArgAspPheArgIleGlnThrPhe          | 313    |
| Db | 992987 | TTGAGAAAGCGTTTCCAAACAATCGGATATCTCTCAGAGATTATCGTGTGCGTAGACCA           | 993046 |
| Qy | 314    | GlySerSerGluLysIleLysIleTyrPheSerAspLysAlaLeuSerTyrThrLysGln          | 333    |
| Db | 993047 | GATAGTTCTGGAAAAGTCAAATATATTATCTCAGAATGCCTTCGATAGAGTTGAGCAA            | 993106 |
| Qy | 334    | IleArg-----AlaSerLeuLeuLys  | 340    |
| Db | 993107 | ATAGAAGGAGCTGGATCAGAACAGAAAGCATCGGATTATCATCTTGTCTCGAGTTCTTCAA         | 993166 |
| Qy | 341    | LeuThrIleMetSerTrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeu          | 360    |
| Db | 993167 | GTATTGTCGTATTC-----   | 993181 |
| Qy | 361    | TyrGlyAsnSerLysArgAsnAlaLysPhe-----TrpSerLysVal                       | 374    |
| Db | 993182 | --GGAAGTTCTACTACTTCTATGTAITTTGATGCTTCTCGAGGTAGCTGGTTTAAGGTA           | 993238 |
| Qy | 375    | SerSerLysLeuSerLysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeu          | 394    |
| Db | 993239 | AGTCGACGAGCTTCCAAATAAATGCGTACACAGGCTGTCGCAATTAATTTGGAGCTTTA           | 993298 |
| Qy | 395    | AlaIleIleLeuLeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaVal          | 414    |
| Db | 993299 | GGAATTATCTTACTTATGTTAGCTTGGCTTTGAGTGGAGATATGCCTTCAGTCGGATT            | 993358 |
| Qy | 415    | CysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPhePhe          | 434    |
| Db | 993359 | TGTGCTTTAATGCACGACTTATTAGCGACTTGTGCTGTAGTTGCTCTTACATTTCIT             | 993418 |
| Qy | 435    | LeuLysLysIle-GlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyr         | 454    |
| Db | 993419 | TTGCAGAGAAATTACAGATTGATTTCAGAGCAATAGGCGCAATTGATGACAGTCTAGGTA          | 993478 |
| Qy | 454    | rSerLeuAsnAsnThrLeuIleIlePheAspArgIleArgGluAspArgGlnAlaAsnLe          | 474    |
| Db | 993479 | CTCTTTGNACCAATACGTTGATTATTTTGTGTCGTATTCGAGGAAGATCGCGCGAAAAACT         | 993538 |
| Qy | 474    | uPheThrProMetHisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrVa          | 494    |
| Db | 993539 | GTTTCAACCCCAATGCGGATTTTGTATGATCAATGATGTCATTGCAGAAAGACATTAGGACGCTACAGT | 993598 |
| Qy | 494    | IleMetThrThrAlaThrLeuSerValLeuLeuMetLeuLeuPheIleGlyLysSerSe           | 514    |
| Db | 993599 | TATGACTACGGCAACAACCTCATCTGATCGTCACTTTTGTATTGTTATTCGTGGGGAGGTT         | 993658 |

| Qy                     | 514   | rValpheasnPhealaphelleMetThrIleGlyleuLeuGlyThrLeuSerSerleu | 534         |
|------------------------|---|--|-------------|
| Db                     | 993659  | TATTTCAAATTTGTCATTATGACATGGGATCTTGTAGGAACGCTATCTTCATT      | 993718      |
| Qy                     | 534   | uTyIlealaproProLeuLeuLeuPheMetValArgLysGluAenArgSer        | 551         |
| Db                     | 993719  | GTACATAGTCTCTCGCTCTTCTTATTTCATGGTGGTAAGGAACAACAAAT         | 993770      |
| RESULT 4               |   |  |             |
| ID                     | AAT42063/c  |  |             |
| XX                     | AAT42063 standard; DNA; 1830121 BP.   |  |             |
| AC                     | AAT42063;   |  |             |
| DT                     | 14-SEP-1999 (first entry)   |  |             |
| DE                     | Haemophilus influenzae complete genome sequence.  |  |             |
| KW                     | Genome; bacterium; Haemophilus influenzae; computer readable medium; expression modulating fragment; regulation; gene expression; vector; organism; open reading frame; ORF; ds.  |  |             |
| OS                     | Haemophilus influenzae.   |  |             |
| PN                     | W09633276-A1.   |  |             |
| PD                     | 24-OCT-1996.  |  |             |
| XX                     | 22-APR-1996; 96WO-US05320.  |  |             |
| PR                     | 07-JUN-1995; 95US-0487429.  |  |             |
| PR                     | 21-APR-1995; 95US-0426787.  |  |             |
| PR                     | 07-JUN-1995; 95US-0476102.  |  |             |
| XX                     | (HUMA-) HUMAN GENOME SCI INC.   |  |             |
| PA                     | (UTJO ) UNIV JOHNS HOPKINS.   |  |             |
| PI                     | Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;   |  |             |
| DR                     | WPI; 1996-485782/48.  |  |             |
| XX                     | Haemophilus influenzae Rd genome recorded on computer readable medium - useful for identifying commercially important nucleic acid fragments by homology searching  |  |             |
| PS                     | Claim 1; Page 77.2-77.1091; 1291pp; English.  |  |             |
| CC                     | This sequence represents the complete genome sequence of the bacterium Haemophilus influenzae strain Rd. The invention relates to a computer readable medium (CRM) having recorded upon it the complete H. influenzae nucleotide sequence (I), a representative fragment of (I) or a nucleotide sequence at least 99% identical to (I). By providing the full-length genomic sequence in a computer readable form, it is possible to identify commercially important nucleic acid fragments and expression modulating fragments (EMFs) of the Haemophilus genome. The EMFs can be used to regulate the expression of a nucleic acid molecule. Vectors and altered organisms comprising the predicted ORFs can be used to produce any of the polypeptide fragments of the H. influenzae Rd genome. |  |             |
| SQ                     | Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;   |  |             |
| Alignment Scores:      |   |  |             |
| Pred. No.:             | 1.78e-65  | Length:  | 1830121     |
| Score:                 | 719.00  | Matches:   | 186         |
| Percent Similarity:    | 55.69%  | Conservative:  | 127         |
| Best Local Similarity: | 33.10%  | Mismatches:  | 174         |
| Query Match:           | 25.88%  | Indels:  | 75          |
| DB:                    | 17  | Gaps:  | 18          |
| US-09-868-987-14       | (1-552) x AAT42063 (1-1830121)  |  |             |
| Qy                     | 13  | AsnHisIaSerValSerGlyLysPhe                                 | -----Thr 22 |

```
Db 268552 AATGTTGCAACAATTCAAGGACGTTTGGTTCTTAATTTCCAAATTTACTGGTGTGTATAGC 268493
Qy 23 HisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheValPro 42
Db 268492 ATTCCGAAGCACATAATCTTTACCTTATTTACCTTATTTAAATCTGGTGCATTAATTCACCAATT 268433
Qy 43 GluValLeuSerGluThrIleSerAspLeuGlyLysGlnCysThrGlnGly 62
Db 268432 CAATATGTTGAGAACCGACANNTGGCCCATCATAGTGGCGAACAACGTAGACAAAGG 268373
Qy 63 IleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyrArg 82
Db 268372 ATTAATCGAGTCTTTGGGATTAAGTTCGTTATTTCCCTTTATGTTGTTTACTACAAA 268313
Qy 83 PheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTyrAlaAla 102
Db 268312 ATGTTTGGTGTGATTGCAAGTTTGGCACTGTTTATTAATATCGTATTACTTGTGGGATTA 268253
Qy 103 LeuGlnTyrLeu---AspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAla 121
Db 268252 ATGTCTATTTTACCCGGCGGACACTTTCAATGCGGGTATTCGGGTATCGTTTAACT 268193
Qy 122 MetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPheLeu 141
Db 268192 TTAGGTATGTCAGTAGATGCAATGTATGATTTTGAACGTATTAAAGAAGAAATTCGT 268133
Qy 142 LeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAlaIle 161
Db 268132 AATGGTCTGTTCAATTACGACGACCATTAATGAAGTTTATTAACGGCGCATTTACTTCTATT 268073
Qy 162 PheAspSerAsnLeuThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGly 181
Db 268072 TTTGATCAACACTTAAACCAACTTTAACCGCAATTTCTATACCGCGGTAGGAACAGGC 268013
Qy 182 ProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThrAla 201
Db 268012 CCAATTCAGGTTTGGATTACGCTTTCACCTGGTGTGCGATTCTATGTTTACCGCG 267953
Qy 202 LeuPheMetThr-----LysPhe 207
Db 267952 ATTACAGGAACCTCGCGCATTTAGTTAATGCCCTTTACGGTGGTAAACAACCTTAAACAAATTA 267893
Qy 208 Phe---PheMetLeuTrpMetAsnLys-----ThrGlnHisThrGlnLeuHisMetMetAs 225
Db 267892 TTAATTTAGCGGGAATGATGAACCTTTTACAAAGATAAGACCGACATTTATTCGG 267833
Qy 225 nLysPheValGlyIleLys-----HisAspPheLeuArgGlyCysLys 240
Db 267832 TGAATCAATGGGATAAAGCTCCCGTTCCTCCATTGACGTGAATTTATGAAGTGGCTAAAT 267773
Qy 240 sLeuTrpAlaValSerGlySerValPheLeuLeuGly-----CysValAlaLeuGly 257
Db 267772 GGGTTATATATTATCCGCACCTTTGTAGTGAATTTCTCTATTTTATTTATTTATACCAAGG 267713
Qy 257 yPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGlyGlyTyrAlaPheThrPh 277
Db 267712 ATTAAAC---TGG-----GGCTTAGATTTTACTGGTGGAGTGTATTGTATAC 267668
Qy 277 eAsnProLysGluHisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLys 297
Db 267667 TCACCTTCGCGAGTCC-----GCTAACCTTGAACAAATTCGTAGT-----AA 267626
Qy 297 sLeuGlnGluAlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGly 317
Db 267625 ACITTCACGAATAATGGAATTTGAAAGCCCA-----ATTGTACAAACACACAGATCGGTCA 267572
Qy 317 uLysIleLysIleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAla 337
Db 267571 GGATGTGATGAT-----CGTTTACTCTGCAAG 267545
Qy 337 rLeuLeuLysLeuThrIleMetSerTrpArgTyrCysGlyIleValValArgAsnArgPr 357
Db 267544 TAATAATGATTCTACCAATT-----GGTGAACACGTCACAAAGATGCTACAGAATGTAGA 267491
```

```
Qy 357 oArgPheLeuTyrGlyAsnSerLysArgAsnAlaLysPheTyrSerLysValSerSerLys 377
Db 267490 TAAAGACATT-----CAAATTCGCAGTATTGAGTTT-----GTTGGCCCAA 267449
Qy 377 sLeuSerLysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaAlaIleIle 397
Db 267448 TGTTGGTGAAGAAATTAGCACAGGTGCGGTATATCGA/CTTTAGCGACATTAGCAATGGT 267389
Qy 397 eLeuLeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLe 417
Db 267388 GCTTATTTATGTTGGGTTCACGTTTGAATGGCGTTTACGCTTTGGCAGTATCGCTTCTCT 267329
Qy 417 uIleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHisPheLeuLysLys 437
Db 267328 TCGGCACGACGTCATTATTATTCGCTAGGGTA-----TTCTTGCATTACA 267284
Qy 437 sIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAs 457
Db 267283 AATTGAATGATCTTACTTTTGTCCGACGATTTTAACTGTGGTGGTACTCCATCAA 267224
Qy 457 nAsnThrLeuIleIlePheAspArgIleArgGluAsp---ArgGlnAlaAsnLeuPheTh 476
Db 267223 CGATAGTATTGGTATTGACCGGGTTCGTGAAATTTCCGAAAAATTTAGACGATTGGA 267164
Qy 476 rProMetHisValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetTh 496
Db 267163 TACGATTGATATTT---ATTGATATTCTTTAACGCAAACTTTATCAAGAACTATCATTA 267107
Qy 496 rThrAlaThrThrLeuSerValLeuLeuMetLeuLeuPheIleGlyLysSerValPh 516
Db 267106 TTCGGTTACTACATAGTTGTCGTATGCGATGTCCTCTCTTTGGTGGTCTTCCATTCA 267047
Qy 516 eAsnPheAlaPheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrI 536
Db 267046 TAACTTTTCACTTGTCTTACTCGTAGTATTGCTAGTTTGTACTTATTCCTCGATTTTGT 266987
Qy 536 eala 537
Db 266986 TGCC 266983
RESULT 5
AAF28535/c
ID AAF28535 standard; DNA; 45613 BP.
XX AAF28535;
AC AAF28535;
XX 04-APR-2001 (first entry)
XX Genomic fragment #22.
XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;
XX bronchopulmonary; endocarditis; meningitis; ss.
XX Moraxella catarrhalis.
OS
XX WO200078968-A2.
XX 28-DEC-2000.
XX 16-JUN-2000; 2000WO-US16649.
XX 18-JUN-1999; 99US-0140121.
XX (INCY-) INCYTE GENOMICS INC.
XX Lagace RE, Patterson C, Berg KL;
XX WPI; 2001-041427/05.
XX Genomic library for identifying diagnostic and therapeutic
XX compositions, and for identifying virulence factors, regulatory
XX elements and drug targets, comprises Moraxella catarrhalis nucleic
PT
```

PT acids -  
 XX PS Claim 1, Page 180-191, 545pp; English.  
 XX  
 CC The present invention relates to a Moraxella catarrhalis genomic library  
 CC comprising of a combination of 41 nucleic acid molecules (see  
 CC AAF28514-AAF28554). The library has a number of uses described in the  
 CC specification e.g. is useful for identifying diagnostic and therapeutic  
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large  
 CC aerobic, gram-negative diplococcus, normally found among the bacterial  
 CC flora of human upper airways. M. catarrhalis is known to cause acute,  
 CC localised infections such as otitis media, sinusitis and bronchopulmonary  
 CC infection and life-threatening, systemic diseases including endocarditis  
 CC and meningitis.  
 XX  
 SQ Sequence 45613 BP; 13249 A; 9346 C; 9947 G; 13071 T; 0 other;

Alignment Scores:  
 Pred. No.: 2,53e-64 Length: 45613  
 Score: 686.50 Matches: 178  
 Percent Similarity: 50.50% Conservative: 126  
 Best Local Similarity: 29.57% Mismatches: 191  
 Query Match: 24.71% Indels: 107  
 DB: 22 Gaps: 14

US-09-868-987-14 (1-552) x AAF28535 (1-45613)

QY 17 ValSerGlyLysPheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGly 36  
 DB AATCCAGGCTTAGATTAGATCGACGAGCGGCTTGTACTGTAGCTCAGGC 28831

QY 37 AlaMetSerPheValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLys 56  
 DB GCGTTAGCGGCACCGCATGCTTCTTGAAGAGCGCAACCATTTGGCCATCTTGGTCAA 28771

QY 57 LysGlnCysThrGlnGlyIleLeuSerAlaCysGlyLeuAlaMetLeuValLeu 76  
 DB GATAATATCGATAAAGGATTATCTCTACACAGGTGGCTATCTGTGTATTTATGG 28711

QY 77 MetSerValTyrrArgPheGlyValIleAlaSerGlyAlaValLeuLeuAsnLeu 96  
 DB ATGATTGTTTTATCGTCATTTGGTGTGTGCAATATGCTTTAGCCCTTTAATGTC 28651

QY 97 LeuLeuIleTrpAlaAlaLeuGlnTyrrLeuAspAlaProLeuThrLeuSerGlyLeuAla 116  
 DB ATTATCATGTTGGCATTATGTCCTTATAGCATCATCGCTCACCTACCTGTTATGCT 28591

QY 117 GlyIleValLeuAlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIle 136  
 DB GGGATTGTTGACCATTTGGTATGCGGTTGATGCAATGTGCTAATTTTGGCGGTATT 28531

QY 137 ArgGluGluPheLeuSerGlnSerLeuLysSerValGluLysGlyTyrrThrLys 156  
 DB CGTGAAGAGCTTGCCAAATGGTACTCGTGCAAAATCCGCCATTTGGCGAGGTTTGCATCGT 28471

QY 157 AlaPheGlyAlaIlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhe 176  
 DB GCATTAGCTCAATTTTGTATGCCAATATCACACGCTATTAAATGCAATATCTATT 28411

QY 177 PheLeuAspThrGlyProIleLysGlyPheAlaLeuThrLeuLeuGlyIlePheSer 196  
 DB GCCATCGGAACAGGACCAATCAAGGTTTGAATCAGCTTGCATATGTTATTCAGT 28351

QY 197 SerMetPheThrAlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThr 216  
 DB TCTTTGTTACTGCGGATTATTAGTCACTCGTGCCATGATTCAAAATTTGGTATGGTCATCGC 28291

QY 217 GlnHisThr-----GlnLeuHisMetMetAsn----- 225  
 DB AAAAACTTAACCAAACTTAGCATCGTTAGGAGTTAAAAATGAATAAATAAATACCCCT 28231

QY 225 ----- 225

DB 28230 ATCGACCAAATGTCAGCACAGCCACCACCGCGTGTGCGCGTGTGATGATGATAA 28171  
 QY 225 ----- 225

DB 28170 GGTGCCAATGTCGCCACAAATCAAGATGCTCTTCCGATCAATCATTTGATGCCGAGCT 28111  
 QY 226 -----LysPheValGlyIleLys-----His 232  
 DB 28110 GGTGATGAAGCTGCGGAAAAAGCGGTGGTGTCAAACTCATCGCTAATCAGCGGTATCAT 28051

QY 233 AspPheLeuArgGlyCysLysLysLeuTrpAlaValSerGlySerValPheLeuLeuGly 252  
 DB CTTTATGAAGCTTGAAGCCCATGATAGTATTTTCGGTATTGCTCTATCTATCATAGG 27991

QY 253 CysValAlaLeuGlyPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGly 272  
 DB CTGATAGCAATCATTTGTCAGGGTTTAAAC-----TTTGGTTGGATTTTACAGGTGCT 27937

QY 273 TyrAlaPheThrPhe-----AsnProLysGluHisGlyIleSerAspValAlaGln 289  
 DB GTGTGACCAATGCTCGCTTATGAACAGCGCTGCTTCAG----- 27901

QY 290 MetArgGlyLysValValHisLysLeuGlnGluAlaGlyLeuSerSerArgAspPheArg 309  
 DB ---CAAGCACAAATCATTAAGCATTTGGCTGCAATGCTTTAAT-----GATCGCGTT 27850

QY 310 IleGlnThrPheGlySerSerGluLysIleLysIleTyrrPheSerAspLysAlaLeuSer 329  
 DB GTCCAAATACCTAGAACTCGCCAAAGAGCTTTGGTAGCTTACCGCCCAAGATGGCTCA 27790

QY 330 TyrThrLysGlnIleArgAlaSerLeuLysLeuThrIleMetSerTrpArgTyrrCys 349  
 DB ---SCTGAAATCTAAGTGCTTCTTTGGATCAG----- 27760

QY 350 GlyIleValValArgAsnArgProArgPheLeuTyrrGlyAsnSerLysArgAsnAlaLys 369  
 DB -----ATTCTTAATTTTCCA-----AATAATCCATCAACCATTCAT 27724

QY 370 PheTrpSerLysValSerSerLysLysLysMetArgTyrrGlnAlaThrIleGly 389  
 DB AGCATTAGCATTTGCGCAGCCAACTGGCAATGAATTTACCTAAATTCATCTGCTCGC 27664

QY 390 LeuLeuGlyAlaLeuAlaIleLeuLeuTyrrValSerLeuArgPheGluTrpGlnTyrr 409  
 DB ATGCTGTGTCATTTGGTATGATGCTCATTTATGTGCAACACGATTTCCAGTTCAGCTG 27604

QY 410 AlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeuPhe 429  
 DB GCTTTGGTGCGTACTTGTCTTTATTTTCATGATG-----ATTATCGTT 27559

QY 430 IleAlaHisPhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMet 449  
 DB GTTGGTATTTTGGCATCATCGGCTGGCCATTTGACTTTGACGGTTTGGCAGCTGTATGT 27499

QY 450 ThrValLeuGlyTyrrSerLeuAsnAsnThrLeuIleIlePheAspArgIleArgGluAsp 469  
 DB GCTTTGATGCTATTCGTCGAATGATACATCGTTGTTATGATCGTATTCGTCGAGAA 27439

QY 470 ---ArgGlnAlaAsnLeuPheThrProMetHisValLeuValAsnAspAlaLeuGlnLys 488  
 DB TTTGCCGAGTGCCTGGTTTAAACACCCCGTCAGGTGCTT---GATTTATCACTACAGAA 27382

QY 489 ThrPheSerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMetLeuLeu 508  
 DB ACCTTAAGACCAACATATGACTGCTGGTAGCGGTGTTGTTGGTTGTTGCTGCTGCTG 27322

QY 509 PheIleGlyLysSerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeu 528  
 DB TTTCTAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 27262

QY 529 GlyThrLeuSerSerLeuTyrrIleAlaPro-----ProLeuLeuLeuPheMetValArg 546  
 DB GGTACATATTTCTCCATTTATGTTGCAAGCTCTATTCGCTGATGATGGTCTATCAAGA 27202

|                        |                    |               |
|------------------------|--------------------|---------------|
| Alignment Scores:      |                    |               |
| Pred. No.:             | 1.8e-65            | Length:       |
| Score:                 | 678.50             | Matches:      |
| Percent Similarity:    | 50.25%             | Conservative: |
| Best Local Similarity: | 29.01%             | Mismatches:   |
| Query Match:           | 24.42%             | Indels:       |
| DB:                    | 20                 | Gaps:         |
| US-09-868-987-14       | (1-552) x AAX29477 | (1-2214)      |

QY 323 pheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeuThr 342  
 Db 1633 TTTAAAGACAAA----- 1644  
 QY 343 IleMetSerTrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeuTyrGly 362  
 Db 1645 -----TACGGA 1650  
 QY 363 AsnSerLysArgAsnAlaLysPheTrpSerLysValSerSerLysLeuSerLysLysMet 382  
 Db 1651 TCTGATCCAAATGTCAGACAGTTTCACCGACAGTCGGTAAGGAGCTGGCAGAAATCGG 1710  
 QY 383 ArgTyrGlnAlaThrIleGlyLeuLeuGlyAlaLeuAlaIleLeuLeuTyrValSer 402  
 Db 1711 CTGTACGCGAGTTGCTATGCTTATGGC-----ATCATTATTACGTTTCA 1758  
 QY 403 LeuArgPheGluTrpGlnTyrAlaPheSerAlaValCysAlaLeuIleHisAspLeuLeu 422  
 Db 1759 ATCCGATTGGAATACAAAATGCGGATTGCTGCCATCGCCTCATTCGTATATGAC----- 1812  
 QY 423 AlaThrCysAlaValLeuPheIleAlaHisPheLeuLysLysIleGlnIleAspLeu 442  
 Db 1813 -----GCATTTCTTATGCTACGTTCTTCAGTATTACAGGCTTGAGGTAGATGTT 1863  
 QY 443 GlnAlaIleGlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIle 462  
 Db 1864 ACATTATCTCGCGCCATCTTGACGATAATCGGTTATTCATTAAAGTACATACTGTACA 1923  
 QY 463 PheAspArgIleArgGlu-----AspArgGlnAlaAsnLeuPheThrProMetHis 479  
 Db 1924 TTTCAGAGGTCGCGGAGCATATGAAAAGCGTAAGCGCGAAACCTTTGCCGATCTGAAC 1983  
 QY 480 ValLeuValAsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThr 499  
 Db 1984 CATATTGTAAACCTGAGCTGCGAGCAACCTTTACAGTTCAATTAACATGTTATTAACC 2043  
 QY 500 ThrLeuSerValLeuLeuMetLeuPheIleGlySerSerValPheAsnPheAla 519  
 Db 2044 GTTGTGATTGTTGTTGTGACATTTGCTGATCTTTGGAGCATCTTCTATTACTAATCTTCA 2103  
 QY 520 PheIleMetThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProPro 539  
 Db 2104 ATTGCTTTATTGCTGGGCTGTTAACAGCGCTTTATTCTTCTATACATTGCCGCACAA 2163  
 QY 540 LeuLeuLeuPheMetValArgLysGluAsnArgSerLys 552  
 Db 2164 ATTTGGCTTGC-ATGGAAGGAGAGAACTGAAAAAGA 2201  
 RESULT 7  
 ID AAX91694  
 XX AAX91694 standard; DNA; 2943 BP.  
 AC AAX91694;  
 XX  
 XX 20-MAR-2003 (updated)  
 DT 25-AUG-1999 (first entry)  
 XX  
 DE Porphyromonas gingivalis protein PG18 encoding DNA.  
 XX  
 KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;  
 KW vaccine; antigenic; ds.  
 XX  
 OS Porphyromonas gingivalis.  
 XX  
 PN WO929870-A1.  
 XX  
 PD 17-JUN-1999.  
 XX  
 PF 10-DEC-1998; 98WO-AU01023.  
 XX  
 PR 10-DEC-1997; 97AU-0000839.  
 PR 31-DEC-1997; 97AU-0001182.  
 PR 30-JAN-1998; 98AU-0001546.

PR 10-MAR-1998; 98AU-0002264.  
 PR 09-APR-1998; 98AU-0002911.  
 PR 23-APR-1998; 98AU-0003128.  
 PR 05-MAY-1998; 98AU-0003338.  
 PR 22-MAY-1998; 98AU-0003654.  
 PR 29-JUL-1998; 98AU-0004917.  
 PR 30-JUL-1998; 98AU-0004963.  
 PR 04-AUG-1998; 98AU-0005028.  
 XX (CSLC-) CSL LTD.  
 XX  
 XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;  
 PI Ross BC, Rachel LJ, Webb EA;  
 XX WPI; 1999-385613/32.  
 DR P-PSDB; AAY34476.  
 XX  
 XX Antigenic Porphyromonas gingivalis peptides for preventing  
 PT gingivitis  
 XX  
 PS Claim 12; Page 202-203; 588pp; English.  
 XX  
 CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic  
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to  
 CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the  
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial  
 CC activity with a vaccine mechanism of action. The PG polypeptides can be  
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can  
 CC be used to detect Porphyromonas gingivalis in standard hybridisation  
 CC assays. Porphyromonas gingivalis is involved in periodontal disease  
 CC especially gingivitis.  
 CC (Updated on 20-MAR-2003 to correct PR field.)  
 XX  
 SQ Sequence 2943 BP; 704 A; 596 C; 719 G; 824 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1,5e-64 Length: 2943  
 Score: 672.00 Matches: 189  
 Percent Similarity: 51.03% Conservative: 109  
 Best Local Similarity: 32.36% Mismatches: 213  
 Query Match: 24.19% Indels: 74  
 DB: Gaps: 18  
 US-09-868-987-14 (1-552) x AAX91694 (1-2943)  
 QY 3 SerSerProIleLeuAsnValProLeuLys---AsnHisAlaSerValSerGlyLysPhe 21  
 Db 1300 TCTGCTCCGAACGTGAATGATGAGATCACGGGCGTCTCAGATCTCCGGGCACCTTC 1359  
 QY 22 ThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheVal 41  
 Db 1360 ACCGTGGAGGCGCGGTGACCTTGCCACGATCACTCCGTAATAATGGATGCTACG 1419  
 QY 42 ProGluValLeuSerGluThrIleSerSerAspLeuGlyLysLysGlnCysThrGln 61  
 Db 1420 GTAAGCATCGAACAGGAAACGTGATTGCTCTACGCTGGGTCCGAGTCCATTAAAGCC 1479  
 QY 62 GlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 81  
 Db 1480 GGATTCCTTGCTGCTCTGCTGCTTTGTTATCTGATGTTTACATGTGCTCGCTTAC 1539  
 QY 82 ArgPhe--GlyGlyValIleAlaSerGlyValValLeuLeuAsnLeuLeuIleTrp 100  
 Db 1540 GGTTCCTTCGCGGTCTTTATCCGAAACGGCGCATGATTGTAACACGCTTCTCACATTG 1599  
 QY 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120  
 Db 1600 GCGTATTGGCTCTTTCCATGCGCTGCTGACCTCTCGGATATCGCAGGTTTGGTCTG 1659  
 QY 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPhe 140  
 Db 1660 ACGTGGGTATGGCTGGATGCCAACGATCTTATCTTCGAGCGTATCAAAAGAGAGCTT 1719

QY 141 LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160  
 DB 1720 CGTCCGGTAAGACTCCGATTCGTCGCTTATGCGAAGCTTCTCTGCC 1779  
 QY 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180  
 DB 1780 ATCTTCGACTCGAAGCTTACGACTATTATTACCGGTATCATCCTTCTCTACGGGACG 1839  
 QY 181 GlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThr 200  
 DB 1840 GGGCGGATTCGGGTTGGCCACTACGTTGATTATCGGTCTATCGCTTCTTTCATTACG 1899  
 QY 201 AlaLeuPheMetThrLysPhePhePhe-MetLeuTrpMetAsnLysThrGlnHisThrG1 220  
 DB 1900 GCTGCTCTTTCGACTCGTATCGTCTTCGAGAACTGGCGAAAAAGGTCGTTGGATAAG 1959  
 QY 220 nLeuHisMetMetAsnLysPheValGlyIleLys-----HisAspPheLeuAr 236  
 DB 1960 ATTACATTCACTACGACGATTAATCTCGCAATCTCTTGTCAATCCCTCATACAAC-ATCTT 2018  
 QY 236 gGlyCysLysLysLeu---TrpAlaValSerGlySerValPheLeuLeuGlyCysValAl 255  
 DB 2019 GGGTAAGCGCAAGACCGCTTATCATCTCCGCTGATTATCATCTGTTTGGACTTATAGC 2078  
 QY 255 a-----LeuGlyPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGlyG1 272  
 DB 2079 TTTCAATTTACAATCGTCTCAATAGG-----GGTATTGAATTTCTCCGAGG 2123  
 QY 272 Y-----TyrAlaPheThrPheAsnProLysGluHisGlyIleSerAspValAlaGlnMe 290  
 DB 2124 ACGTAACTACGTAGTAAATTC-----GACCAGCCTGTATCTTCCGAAGCC---GT 2171  
 QY 290 tArgGlyLysValValHisLysLeuGlnGlu-----AlaGlyLeuSe 304  
 DB 2172 TCGTTCGGCTTGCTTCTCCCTTCAGGAAAGGATTGGTTACCTCCATCGGTACTGA 2231  
 QY 304 rSerArgAspPheArgIleGlnThr-----PheGlySerSerGlyLysIleLy 320  
 DB 2232 AGGACAGAGGTGGTATATCTACGAATAAGTATAGATCCAGGAGGAAAGCGAAGAACTGA 2291  
 QY 320 sIleTyPheSerAspLysAlaLeuSer-----TyrThrLysGlnIleAr 335  
 DB 2292 AGCAGAGATTACTGACAAATTTGTATCAGAGCTGAAAGGTTTCTACACCCAGCAGCCTAC 2351  
 QY 335 gAlaSerLeuLeuLysLeuThrIleMetSerTrpArgTyCysGlyIleValValArgAs 355  
 DB 2352 TGCTGATCAGTCTTGACAAATATCATAGCTCTCAG----- 2388  
 QY 355 nArgProArgPheLeuTyArgLysSerLysArgAsnAlaLysPheTrpSerLysValSe 375  
 DB 2389 -----AAAGTAAG 2396  
 QY 375 rSerLysLeuSerLysLysMetArgTyGlnAlaThrIleGlyLeuLeuGlyAlaLeuAl 395  
 DB 2397 TCCAGTATGTCGAGTGCACATCCAGAGAGGTGCTATTGGGCTGTGCTTATCGATGAT 2456  
 QY 395 alIleLeuLeuTyValSerLeuArgPheGluTrpGlnTyAlaPheSerAla----- 413  
 DB 2457 CTTCATGCCCATTTATATCTGATTCGCTCCGT---GACATTTCTTCTCTCCCGGGT 2513  
 QY 414 -----ValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeuPhe-- 429  
 DB 2514 ATTCGTATCTGTGGCCGCTACTACATTCGCAATTATTGCTCTGTATCGGTGTGTGGAA 2573  
 QY 430 -IleAlaHisPheLeuLysLysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMe 449  
 DB 2574 GATTCTCCCTTC-----ACCATGGAGATCGATCAGAACTTATCTCGCTCTATTCT 2624  
 QY 449 tThrValLeuGlyTySerLeuAsnAsnThrLeuIleIlePheAspArgIleArgGluAs 469  
 DB 2625 GGCTATCATCGGTACTCGCTCAATGACACCGGTGGTGTATTGACCGTATCCGAGAGAC 2684  
 QY 469 pArgGlnAlaAsnLeuPheThrProMetHisValLeuValAsnAspAlaLeuGlnLysTh 489

DB 2685 GATGAAATTTGATACCCCAACAGAGATCGCTATCATAGGTGATCAACGATGCCCTTAATCAAC 2744  
 QY 489 rPheSerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMetLeuLeuPh 509  
 DB 2745 ATTGGGTGCAACATTAATAACGCTTTTGACTACGTTTATCGTTATGTTGTAATCTTCAT 2804  
 QY 509 eIleGlyGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeuG1 529  
 DB 2805 CTTTGAGGTGCTACGATCGGTAGTTTCAGTTCTCGATCCTCGGTATCGTTATCGG 2864  
 QY 529 yThrLeuSerSerLeuTyIleAlaProProLeuLeuLeuPheMetValArgLysGluAs 549  
 DB 2865 TACATACTCTACGCTCTTTGTTGCTACACCC---CTTGCTACGAGATCCAAAGCGCAA 2921  
 QY 549 nArgSerLys 552  
 DB 2922 GCTCAACAAA 2931  
 RESULT 8  
 AAX91570  
 ID AAX91570 standard; DNA; 2955 BP.  
 XX AAX91570;  
 AC AAX91570;  
 XX XX  
 DT 20-MAR-2003 (updated)  
 DT 25-AUG-1999 (first entry)  
 XX XX  
 DE Porphyromonas gingivalis protein PG18 ORF encoding DNA.  
 XX XX  
 KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;  
 KW vaccine; antigenic; ds.  
 XX XX  
 OS Porphyromonas gingivalis.  
 XX XX  
 PN W09929870-A1.  
 XX XX  
 PD 17-JUN-1999.  
 XX XX  
 PF 10-DEC-1998; 98WO-AU01023.  
 XX XX  
 PR 10-DEC-1997; 97AU-0000839.  
 PR 31-DEC-1997; 97AU-0001182.  
 PR 30-JAN-1998; 98AU-0001546.  
 PR 10-MAR-1998; 98AU-0002264.  
 PR 09-APR-1998; 98AU-0002911.  
 PR 23-APR-1998; 98AU-0003128.  
 PR 05-MAY-1998; 98AU-0003338.  
 PR 22-MAY-1998; 98AU-0003654.  
 PR 29-JUL-1998; 98AU-0004917.  
 PR 30-JUL-1998; 98AU-0004963.  
 PR 04-AUG-1998; 98AU-0005028.  
 XX XX  
 (CSLC-) CSL LTD.  
 XX XX  
 PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;  
 PI Ross BC, Rothel LJ, Webb EA;  
 XX XX  
 DR WPI; 1999-385613/32.  
 DR P-PSDB; AAY34352.  
 XX XX  
 PT Antigenic Porphyromonas gingivalis peptides for preventing  
 PT gingivitis  
 XX XX  
 PS Claim 12; Page 118; 588pp; English.  
 CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic  
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to  
 CC AAY34583; AAX91802 to AAX91989 represent PCR primers used in the  
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial  
 CC activity with a vaccine mechanism of action. The PG polypeptides can be  
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can  
 CC be used to detect Porphyromonas gingivalis in standard hybridisation



CC assays. Porphyromonas gingivalis is involved in periodontal disease especially gingivitis.  
 CC (Updated on 20-MAR-2003 to correct PR field.)  
 XX

SQ Sequence 2955 BP; 711 A; 698 C; 721 G; 825 T; 0 other;

## Alignment Scores:

Pred. No.: 1, 51e-64 Length: 2955  
 Score: 672.00 Matches: 189  
 Percent Similarity: 51.03% Conservative: 109  
 Best Local Similarity: 32.36% Mismatches: 213  
 Query Match: 24.13% Indels: 74  
 DB: 20 Gaps: 18

US-09-868-987-14 (1-552) x AAX91570 (1-2955)

```

QY 3 SerSerProLeuLeuAsnValProLeuLys---AsnHisAlaSerValSerGlyLysPhe 21
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1312 TCTGCTCCGAACGTAATGATGAGATCAGCGCGGTCTCTCAGATCTCCGGGCACTTC 1371
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 22 ThrHisAGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheVal 41
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1372 ACCGTGGAGAGCGCGGTACCTTCCCAACGTACTCAACTCCGTAATAATGATGCTACG 1431
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 42 ProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysGlnCysThrGln 61
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1432 GTAACATCGAACAACGATGATGGTCTACGCTGGTGGCGAGTCCATTAAGCC 1491
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 62 GlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyr 81
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1492 GGAATCTTGTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1551
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 82 ArgPhe---GlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrp 100
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1552 GTTTCCTTCCCGGCTTATCGCAACCGCGCATGATGTAACCGAGCTTCTTCACATG 1611
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1612 GCGGATTTGGCTCTTCCATGCGGTGCTGACCTCTCGGTATCGCAGGTTTGGTGTG 1671
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPhe 140
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1672 ACGCTGGGTATGGCTGTGGTGCACCACTTATCTTCCGCGGTATCAAAAGAGAGCTT 1731
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 141 LeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGlyVal 160
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1732 CGTCCGGTAGACCGGATTCGTCGGTACGATGTTATGGCAACGCTTTCTCTGCC 1791
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1792 ATCTTCGACTCGAACGTTACGACTATTATTACCGGTATCATCCTATTCTCTACGGGAG 1851
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 181 GlyProIleLysGlyPheAlaLeuThrIleLeuLeuGlyIlePheSerSerMetPheThr 200
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1852 GGGCGGATTCGGGTTTGGCCACTACGTATGATTCGCTTATCGCTTCTTCTTATTACG 1911
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 201 AlaLeuMetThrLysPhePhePhe-MetLeuTrpMetAsnLysThrGlnHisThrG1 220
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1912 GCTGCTCTTTCGACTCGTATGCTTCGAGAACTGGCGAAAAGGTCGTTTGGATAAG 1971
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 220 nLeuHisMetMetAsnLysPheValGlyIleLys-----HisAspPheLeuAr 236
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 1972 ATTACATTCACTACGAGCATTTACTCGCAATCTCTTGTCTCAATCCCTCATACAC-ATCTT 2030
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 236 gGlyCysLysLysLeu---TrpAlaValSerGlySerValPheLeuLeuGlyCysValAl 255
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2031 GGGTAAGCGCAAGACCGGCTTATCATTCGGGTGATTTATCATCGTTTGGGACTTATAGC 2090
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 255 a-----LeuGlyPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGlyG1 272
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2091 TTCAATTAACAATCGGCTCTCAATAGG-----GGTATTGAATCTCCGGAGG 2135
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 272 y-----TyrAlaPheThrPheAsnProLysGluHisGlyIleSerAspValAlaGlnMe 290

```

RESULT 9

ABK74549

ID ABK74549 standard; DNA; 1584 BP.

XX

AC ABK74549;

XX

DT 13-AUG-2002 (first entry)

XX

DE Bacillus licheniformis genomic sequence tag (GST) #1840.

XX

KW Differential gene expression; genomic sequenced tag; GST;  
 KW altered culture condition; environmental stress;  
 KW physiological provocation; ds.

OS Bacillus licheniformis.

PN WO200229113-A2.

XX 11-APR-2002.

XX 05-OCT-2001; 2001WO-US31437.

XX 06-OCT-2000; 2000US-0680598.

XX 27-MAR-2001; 2001US-279526P.

XX (NOVO ) NOVOZYMES BIOTECH INC.

XX (NOVO ) NOVOZYMES AS.

XX Berka R, Clausen IG;

XX WPI; 2002-416684/44.

XX Monitoring differential expression of several genes in first Bacillus  
 XX cell relative to expression of same genes in one or more second  
 XX Bacillus cells, by using substrate containing Bacillus genomic  
 XX sequenced tag array

PS Claim 4; SEQ ID NO 1840; 200pp; English.

XX The invention describes a method of monitoring differential expression of  
 CC genes in a first Bacillus cell relative to expression of the genes in  
 CC other Bacillus cells, comprising hybridising labelled nucleic acid probes  
 CC isolated from Bacillus cells to a substrate containing array of Bacillus  
 CC genomic sequenced tags (GST), examining the array, and determining  
 CC relative gene expression by an observed hybridisation reporter signal of  
 CC a spot in the array. The method is useful for measuring the expression of  
 CC genes in a first Bacillus cell relative to expression of the same genes  
 CC in one or more second Bacillus cells. The method is useful for monitoring  
 CC global expression of several genes from a Bacillus cell, discovering new  
 CC genes, identifying possible functions of unknown open reading frames and  
 CC monitoring gene copy number variation and stability. Monitoring changes  
 CC in expression of genes may be used to provide a representation of the way  
 CC in which Bacillus cells adapt to changes in culture conditions.  
 CC environmental stress or other physiological provocation. Extensive  
 CC follow-up characterisation is unnecessary, when one spot on an array  
 CC equals one gene or one open reading frame, since sequence information is  
 CC available. This sequence represents a genomic sequence tag (GST) used in  
 CC the method of the invention.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1584 BP; 389 A; 347 C; 395 G; 453 T; 0 other;

#### Alignment Scores:

| Pred. No.:             | 4,53e-64 | Length:       | 1584 |
|------------------------|----------|---------------|------|
| Score:                 | 664.00   | Matches:      | 173  |
| Percent Similarity:    | 48.98%   | Conservative: | 114  |
| Best Local Similarity: | 29.52%   | Mismatches:   | 177  |
| Query Match:           | 23.90%   | Indels:       | 122  |
| DB:                    | 24       | Gaps:         | 16   |

US-09-868-987-14 (1-552) x ABK74549 (1-1584)

Qy 17 ValSerGlyLysPheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGly 36  
 Db 21 ATTCAGGCACTTTACGATTACGAGCAAGCAACATCGGAGCATTTTAAATCGCGGA 80  
 Qy 37 AlaMetSerPheValProGluValLeuSerGluGlu-----ThrIleSerSerAsp 53  
 Db 81 GCTCTT-----CCTGTCAAACTGTTGAAAGATTTCGACATCCGTCGGAGCGCAG 131

Qy 54 LeuGlyLysLysGlnCysThrGlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeu 73  
 Db 132 TTGGGAGCAGCGTTTGAATGATACCGGTGTTGGGAATCATTTGTTATGCTATTATT 191  
 Qy 74 IleValLeuMetSerValTyrTyrArgPheGlyValIleAlaSerGlyAlaValLeu 93  
 Db 192 TTCTATTATGCTTCTTACTATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 242  
 Qy 94 LeuAsnLeuLeuIleTrpAlaLeuGln-----TyrLeuAspAlaProLeu 110  
 Db 243 AGCTTTTCGGTCTACATCTACATTACATGCGAGATATTCGACTGGGATGCGGTCTG 302  
 Qy 111 ThrLeuSerGlyLeuAlaGlyIleValLeuAlaMetGlyMetAlaValAsnVal 130  
 Db 303 ACCCTTCGGGATCGGGCTCTTATATGGCGTCGCGATGGCGTTGACGCCACATC 362  
 Qy 131 LeuValPheGluArgIleArgGluGluPheLeuLeuLeuSerGlnSerLeuLysSerVal 150  
 Db 363 ATTACGTATGAACGCATAAAGAGAGCTGAAGCTCGGCAATCGGTCGCTGCTGCTGCT 422  
 Qy 151 GluLysGlyTyrThrLysAlaPheGlyAlaIlePheAspSerAsnLeuThrThrValLeu 170  
 Db 423 AAAGCAGGAAACAGAGATCGTTTGCACGATTTTGTATGCCAATATTCAGCAGATGCTT 482  
 Qy 171 AlaSerAlaLeuLeuPhePheLeuAspThrGlyProIleLysGlyPheAlaLeuThrLeu 190  
 Db 483 GCGGNAATCGTCTTTTTCATATTTGGTACAAGCTCTGTCAAGGCTTCGCGACCATGCTG 542  
 Qy 191 IleLeuGlyIlePheSerSerMetPheThrAlaLeuPheMetThrLysPhePheMet 210  
 Db 543 ATCTATCATTTTGCACAGCTTTTATACAGCTGCTCTCTCGCGCTTCTGCTCGGC 602  
 Qy 211 Leu-----TrpMetAsnLysThrGln----- 217  
 Db 603 CTGCTGTCGAAAGCGCTGCTTGCAGGAAAAAGGCTGTTGGGTGCGCAAAAAA 662  
 Qy 218 -----HisThrGlnLeuHisMetMetAsnLysPheValGly 229  
 Db 663 GACATTTTGGATATCAGAAAGACAGCAAGAAATACGGAGCTCCGNAACCGTT----- 716  
 Qy 230 IleLysHisAspPheLeuArgGlyCysLysLysLeuTrpAlaValSerGlySerValPhe 249  
 Db 717 TCCAAATGGATTTCGTCGGAAGCGCAAAATGGTTCTTTGTCATTTTCGGCGCTTCTCTG 776  
 Qy 250 LeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerValLeuGlyMetAspPhe 269  
 Db 777 GCGCTGGTCTTATCGTCTTGTGTTAAGCTGAAC-----CTCGCATCGATTTC 830  
 Qy 270 LysGlyGlyTyrAlaPheThrPheAsnProLysGluHisGlyIleSerAspValAlaGln 289  
 Db 831 TCAAGCGGA-----TCGACA 845  
 Qy 290 MetArgGlyLysValValHisLysLeu-----GlnGlu 300  
 Db 846 ATCGAAGTTCAAAGCGCATCATAGCTGACACCCAGCAGCTCGAAAGGACTTTGAACAA 905  
 Qy 301 AlaGlyLeu-----SerSerArgAspPheArgIleGln 311  
 Db 906 GTCGGGCTTTGATCTCTGACTCGATCGTCTTTTCAGCGCAAAAAAATCATCGCGCGCC 965  
 Qy 312 ThrPhe-----GlySerSerGluLysIleLysIleTyrPheSerAsp 325  
 Db 966 AGTTTTCGCGCTGCCGATCAGAAAAAGATCGCTGAAGTGAAGATTATTTTAAAGAC 1025  
 Qy 326 LysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeuThrIleMetSer 345  
 Db 1026 AAA----- 1028  
 Qy 346 TrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeuTyrClyAsnSerLys 365  
 Db 1029 -----TACGGAAGCGCAACCA 1043  
 Qy 366 ArgAsnAlaLysPheTrpSerLysValSerSerLysLeuSerLysLysMetArgTyrGln 385

```

Db 1044 AATGTCAGCACCGTATCGCTACCGTCGCGAAGAGCTGGCGAGAAATGCGCTTTATCGG 1103
Qy 386 AlathrileGlyLeuLeuGlyAlaLeuAlaLeuLeuLeuLeuValSerLeuArgPhe 405
Db 1104 GTGATCATCGCGTCAATCGGG-----ATCATTTTGTATGTCGATCGCGCTTT 1151
Qy 406 GluTrpGlnTyrAlaPheSerAlaValCysAlaLeuLeuHisAspLeuLeuAlaThrCys 425
Db 1152 GAATACAAGATGGCGATTGCGGCTATTACTTCAATTCGCTGATGAC----- 1196
Qy 426 AlaValLeuPheLeuAlaHisPheLeuLeuLysIleGlnIleAspLeuGlnAlaIle 445
Db 1197 GCGTATTTTCATTCATTCGCGCTTCAGCGCTGCAAGCGTTGAGGTGATGTTACTTTCATT 1256
Qy 446 GlyAlaLeuMetThrValLeuGlyTyrSerLeuAsnAsnThrLeuLeuIlePheAspArg 465
Db 1257 GCTCGCGTGTGAGATCATCGGCTATTTCGATAATGATACAATTTGTAACGTTTGACAGG 1316
Qy 466 IleArgGlu-----AspArgGlnAlaAsnLeuPheThrProMetHisValLeuVal 482
Db 1317 ATCAGGAGCATATCAAAAACGCAACCGAAACCTTCAGCGACTTGTCCATATTGTG 1376
Qy 483 AsnAspAlaLeuGlnLysThrPheSerArgThrValMetThrThrAlaThrThrLeuSer 502
Db 1377 AATTGAGCTTGCAGCAGACCTTCACGCGTTTCGATTAAATACGCTGCTGCTGCTCATT 1436
Qy 503 ValLeuLeuMetLeuLeuPheIleGlyCysSerSerValPheAsnPheAlaPheIleMet 522
Db 1437 GTCGTGATTCACATTCCTCATCTTCGCGCGCTCGATTCGAACTTCTCTGTTCGCGTG 1496
Qy 523 ThrIleGlyIleLeuLeuGlyThrLeuSerSerLeuTyrIleAlaProProLeuLeuLeu 542
Db 1497 TTGTCGGACTTTTGAGCGGAGTATATTCATCATCTCTACATCGCAGCTCAACTATGCTT 1556
Qy 543 PheMetValArgLysGlu 548
Db 1557 GTCTGGAAAGGCGCGAA 1574

RESULT 10
ABT15148
ID ABT15148 standard; DNA; 2292 BP.
XX AC ABT15148;
XX AC
XX DT 06-MAR-2003 (first entry)
XX DE Pathogen specific antigen related staphylococcal DNA SEQ ID No 534.
XX KW Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;
XX KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;
XX KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;
XX KW autoimmune disease; HIV; hepatitis; gene; ds.
XX OS Staphylococcus sp.
XX FN WO200259148-A2.
XX PD 01-AUG-2002.
XX PF 21-JAN-2002; 2002WO-EP00546.
XX PR 26-JAN-2001; 2001AT-0000130.
XX PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX PA Meinke A, Nagy E, Von Ahse U, Klade C, Henics T, Zauner W;
XX PI Minh DB, Vytvytska O, Etz H, Dryla A, Weichhart T, Hafner M;
XX PI Tempelmaier B;
XX DR WPI; 2003-075410/07.
XX PT Identifying, isolating and producing hyperimmune serum-reactive

```

```

PT antigens from a pathogen, for preparing vaccine or medicament for
PT treating or preventing e.g. staphylococcal infections, comprises
PT providing antibody preparation -
XX Example 7; Page 245; 252pp; English.
PS The invention relates to a novel method for identifying, isolating and
XX producing hyperimmune serum-reactive antigens from a pathogen, tumour,
CC allergen, a tissue or host prone to auto-immunity, where the antigens,
CC antigen, a tissue or host prone to auto-immunity, where the antigens,
CC are used in a vaccine, comprises providing antibody preparation from a
CC plasma pool of a type of animal, or individual sera with antibodies
CC against the specific pathogen, tumour, allergen, tissue or host prone to
CC auto-immunity. The hyperimmune serum-reactive antigens comprising any of
CC the 62 sequences of 53-2261 amino acids fully defined in the
CC specification, or their hyperimmune fragments are useful for the
CC manufacture of a pharmaceutical preparation, particularly a vaccine
CC against staphylococcal infections or colonisation against S. aureus or S.
CC epidermidis. The preparation of antibodies is useful for the manufacture
CC of a medicament for treating or preventing staphylococcal infections or
CC colonisation against S. aureus or S. epidermidis. The antibody
CC preparations may also be used for diagnostic and imaging purposes. Other
CC conditions that can be treated include cancer, autoimmune diseases or
CC infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or
CC protozoan pathogens. This polynucleotide sequence represents
CC staphylococcal DNA relating to the method for identifying and producing
CC pathogen specific antigens of the invention.
XX
SQ Sequence 2292 BP; 817 A; 332 C; 400 G; 739 T; 4 other;

```

## Alignment Scores:

```

Pred. No.: 1-39e-62 Length: 2292
Score: 653.00 Matches: 166
Percent Similarity: 51.31% Conservative: 128
Best Local Similarity: 28.97% Mismatches: 199
Query Match: 23.51% Indels: 80
DB: 25 Caps: 13

```

US-09-868-987-14 (1-552) x ABT15148 (1-2292)

```

Qy 2 ValSerSerProLeuLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLysPhe 21
Db 616 GTAGACCAACCTATTAACTTAGTGTGTTGAAATTTTCAGGTGGCTTTCAGGAAAAA 675
Qy 22 ThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPhe--- 40
Db 676 GGTGTTGAAGACGCAACAAATAGCTGAGTTTAAATGCCGCTCATACCAGTTGAT 735
Qy 41 ValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysLysGlnCysThr 60
Db 736 TTAAGAAGAAATTTACTTAACTCWNRDGTTGGTGCAACAATTTGGTCAAGATGCTCTT 795
Qy 61 GlnGlyIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValYtr 80
Db 796 AAGACCATGTTGCTCAATTTAGGTATAGCAATTAATTTATTTATTTATGCTTGTTC 855
Qy 81 TyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrp 100
Db 856 TATCGTTTGGCTGTTAGTTGCAATCATTCATTCCTTAACCTATATATTTAACTTTA 915
Qy 101 AlaAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeu 120
Db 916 GTCGCATTTCAATTTTCATATCAGGTGATTAACCTACCTGATGGCGGATAGTTT 975
Qy 121 AlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluPhe 140
Db 976 GGTGTAGGTATGCTGTCGATGCAATATCATATATGTAACGATTAATAAGATGAAC 1035
Qy 141 LeuLeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAla 160
Db 1036 AGAATTGGACGCGCTTAAACAAAGCGTATTCAAAAGCAAAATAAAGTTTCACTT 1095
Qy 161 IlePheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThr 180

```

Db 1096 ATATTGATTCCTCAACTTAACAACGTGTCATCGTCAGCTGTGCTTTCTTCTTTGGAGAA 1155  
 Qy 181 GlyProlLeuLysGlyPheAlaLeuThrLeuLeuLeuGlyllePheSerSerMetPheThr 200  
 Db 1156 AGTTCAGTCAAGAGCGTTCGCAACCATGTACTCTTAGGTATTTTAAATGATATTTGTAACC 1215  
 Qy 201 AlaLeuPheMetThrLysPhePheMetLeu----- 211  
 Db 1216 GCAGTATTCCTATCAAGAGGGTGTGTATCATCTACTGTGTATCTTCAAACTCTCTTAAAAAA 1275  
 Qy 212 -----TrpMetAsnLysThrGlnHisThrGlnLeuHisMetAsnLysPheValGly 229  
 Db 1276 CAATACCTGGTTATTGGTTGTTAAGCAAGAGATAGACATGATATTAATGAAGGTAAGAT 1335  
 Qy 230 IleLysHisAsp-----PheLeuArgGlyCysLysLys 240  
 Db 1336 GTA---CATGATTTTAAAAACATCATATCAAAAGGTAAACCTTTGTTAAATTAGCTAAGCCA 1392  
 Qy 241 LeuTrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAla 260  
 Db 1393 CTTATTCACCTAGTATTTAAATTTGTAATTTGTTGTTGTTGTTATCATTTCAATATTAAAA 1452  
 Qy 261 TrpAsnSerValLeuGlyMetAspPheLysGlyGlyTyrAlaPheThrPheAsnProLys 280  
 Db 1453 TTAAC-----TTAGGTATTGATTTCTCATCCGGAACAAGCAGCATATTCAATCTAAA 1506  
 Qy 281 GluHisGlyLysSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGlu 300  
 Db 1507 AAT-----GCTATAACAACACAGCAGCTGTGAGAAAACTGTAAAAATCA 1548  
 Qy 301 AlaGlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLys 320  
 Db 1549 GTTGATTTGACACAGAT-----CAATACAGATTAAATGCTAGTGGAAATAAA----- 1596  
 Qy 321 IleTyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLys 340  
 Db 1597 -----AATGCCACAGTTCAGTTTAAAAAAGATTTATCACGT 1632  
 Qy 341 LeuThrIleMetSerTrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeu 360  
 Db 1632 ----- 1632  
 Qy 361 TyrGlyAsnSerLysArgAsnAlaLysPheTrpSerLys----- 373  
 Db 1633 GAGGAACACATAAATAAGTCTGAAGTGAATCTGAATTTGGAGATAATCCACAAT 1692  
 Qy 374 -----ValSerSerLysLeuSerLysMetArgTyrGlnAlaThrIleGlyLeuLeu 391  
 Db 1693 AATACCGTTTCACCTCTCATAGGCCAAGAGCTAGCTAAATAATGCTGTAACGTCAATA 1752  
 Qy 392 GlyAlaLeuAlaIleLeuLeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPhe 411  
 Db 1753 CTTGCTTCTATAGGCATATTATCTATGTTTCACTAAGATTGCAATGCGTATGGGTCTA 1812  
 Qy 412 SerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAla 431  
 Db 1813 TCTTCTGACTGCTGATTTATACATGAGTATTT-----ATCATCATTTGCA 1857  
 Qy 432 HisPhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrVal 451  
 Db 1858 ATCTTTAGTTGTTTACAGTAGAGTAGATTAACTTTATTGTCAGCAGTATTAACTATC 1917  
 Qy 452 LeuGlyTyrSerLeuAsnThrLeuIlePheAspArgIleArgGluAsp---Arg 470  
 Db 1918 GTTGGTTATTCAATCAATGATACAAATCGTAACCTTTCACCGCTGTTCGAGAAAAATCTGCAT 1977  
 Qy 471 GlnAlaAsnLeuPheThr-----ProMetHisValLeuValAsnAspAlaLeuGln 487  
 Db 1978 AAGTTAAAGTAATTACGCATCTACTGATCAAAATGATGATATAGTCAACCGCTCTATTAGA 2037  
 Qy 488 LysThrPheSerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMetLeu 507  
 Db 2038 CAAATATGACAGCTTCTATTATAACAGGTGTGACTGTAGTTGTAGTTGTAGTTGCAATA 2097

Qy 508 LeuPheIleGlyGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeu 527  
 Db 2098 TTAATATTAGGTGACCAACAATATTAAATTTCTCTTTAGCATTTACTAATTTGATTATTA 2157  
 Qy 528 LeuGlyThrLeuSerSerLeuTyrIleAlaProProLeu 540  
 Db 2158 TCTGGTGTATTCTCGTCAATTTTCATTTGCTGTACCATTA 2196

## RESULT 11

AAH52892  
ID AAH52892 standard; DNA; 2226 BP.

XX AAH52892;  
AC

XX 03-SEP-2001 (first entry)  
DT

DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:1177.  
XX

XX Staphylococcus epidermidis SR1 strain; infection; diagnosis;  
KW vaccination; endocarditis; ds.  
XX

OS Staphylococcus epidermidis.  
XX

XX WO200134809-A2.  
PN

XX 17-MAY-2001.  
PD

XX 09-NOV-2000; 2000WO-US30782.  
PF

XX 09-NOV-1999; 99US-0164258.  
PP

XX (GLAX ) GLAXO GROUP LTD.  
PA

XX Kimmerly WJ;  
PI

XX WPI; 2001-316495/33.  
XX P-PSDB; AAG82042.  
DR

XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis -  
XX  
XX Claim 8; Page 337-338; 2188pp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
CC (I) and (II) can have antibacterial activity and therefore can be used  
CC in vaccination. The nucleic acids (I) may be used to produce the  
CC S. epidermidis polypeptides (II) via the production of vectors  
CC containing them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH5090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH53091 to  
CC AAH5098 represent oligonucleotide sequences and primers which are used  
CC in the exemplification of the present invention.

CC N.B. The present invention specifically claims all the polynucleotide  
CC sequences given in the sequence listing of the present specification,  
CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 2226 BP; 796 A; 320 C; 387 G; 723 T; 0 other;  
SQ

## Alignment Scores:

Pred. No.: 7,116-62 Length: 2226  
 Score: 646.50 Matches: 164  
 Percent Similarity: 51.22% Conservative: 129  
 Best Local Similarity: 28.67% Mismatches: 200  
 Query Match: 23.27% Indels: 79

```

DB:                22          Gaps:          12
US-09-868-987-14 (1-552) x AAHS2892 (1-2226)

QY      2 ValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLysPhe 21
DB      550 GTAGACCAACCTATTAAATCTAGTAGTGTGAAATTTTCAGGTGGCTTCAATGGGAAAAA 609

QY      22 ThrHisArgGlyValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheVal 41
DB      610 GGTGTGTGAAGAACCAACATAGCTGAGTTATTAAATGCCGCTCATACCAGTTGAT 669

QY      42 ProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysGlnCysThrGln 61
DB      670 TTAAGAGAAATTTACTCTAACTCTGTGTGCACAAATTTGGTCAAGATGCTCTTGTATAG 729

QY      62 GlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyrTrp 81
DB      730 ACCATGTTTGCATCAATTTGATGATAGCATTAATTTATTATTATCTGGTTCTTAT 789

QY      82 ArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrpAla 101
DB      790 CGTTTGCCTGGTTAGTTAGTGCATCATTCGCTTAACCACTTATATTATTAACTTAGTC 849

QY      102 AlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAla 121
DB      850 GCATTCAATTTTCATATCAGGTGATTAATCTCTACCTGCGATTGGCGCATTTAGTTTAGGT 909

QY      122 MetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPheLeu 141
DB      910 GTAGTATGGCTGTCGATGCCAATATCATATGATGAACGTTATTAAGATGAACATAAGA 969

QY      142 LeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGlyAlaIle 161
DB      970 ATTGACCGCAGCTTAACAAGCGTATTCAAAAGCAAAATAAAGTTCAATCTTAACATA 1029

QY      162 PheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGly 181
DB      1030 TTTGATTCCAACTTAACAACGTGTCGTCGAGCTGTCTTTCTTCTTTGGAGAAAGT 1089

QY      182 ProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThrAla 201
DB      1090 TCAGTCAAGGCTTCGCCAACCACTGTTACTCTTAGGTATTTTAATGATATTTGTAAACGCA 1149

QY      202 LeuPheMetThrLysPhePheMetLeu
DB      1150 GTATTCTTATCAAGAGGTTGTTATCATTTACTGGTATCTTCAAACTTCTTAAAAAACAA 1209

QY      212 ---TrpMetAsnLysThrGlnHisThrGlnLeuHisMetMetAsnLysPheValGlyIle 230
DB      1210 TACTGGTTATTTGGTGTAAAGAAGAGGATAGACATGATATTAATGAAGGTAAGATGTA 1269

QY      231 LysHisAsp
DB      1270 ---CATGATTTAAACATCATATGAAAGGTTAAACTTTGTAAATTAAGTAAAGCCACTT 1326

QY      242 TrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrp 261
DB      1327 ATTTCACTTAGTATTTAATGTAATTTATGTTGTTGATTATCATTTCAATATTAAATTA 1386

QY      262 AsnSerValLeuGlyMetAspPheLysGlyGlyTyrAlaPheThrPheAsnProLysGlu 281
DB      1387 AAC-----TTAGGTATTGATTCTCATCCGGAACAGCAGGATATTCAATCTAAATAAT 1440

QY      282 HisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGluAla 301
DB      1441 -----GCTATAACAACAGCAGGTTGAGAAACTGTAAATACAGTT 1482

QY      302 GlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLysIle 321
DB      1483 GGATTTGGAAACAGAT-----CAAAATACAGATTAATGGTAGTGGAAATAA----- 1527

QY      322 TyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLeuLysLeu 341

```

```

DB      1528 -----AATGCCACAGTTTCAGTTTAAAAAAGATTTTATCAGCT--- 1563
QY      342 ThrIleMetSerTrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeuTyr 361
DB      1564 -----GAG 1566

QY      362 GlyAsnSerLysArgAsnAlaLysPheTrpSerLys
DB      1567 GAACACATAAATAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 373

QY      374 ---ValSerSerLysLysSerLysMetArgTyrGlnAlaThrIleGlyLeuLeuGly 392
DB      1627 ACCGTTTCACTCTCATAGGCAAGAGTAGTAAAAAATGCTGAACATGCAATTAATCTT 1686

QY      393 AlaLeuAlaIleIleLeuLeuTyrValSerLeuArgPheGluTyrGlnTyrAlaPheSer 412
DB      1687 GCTTCTTAGGCAATTTATTTATCTATGTTCAATGATTTGAAATGGCGTATGGGTCTATCT 1746

QY      413 AlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHis 432
DB      1747 TCTGTACTTGCATTTATACATGACGTATTT-----ATCATCATTTGCAATC 1791

QY      433 PhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeu 452
DB      1792 TTTAGTTTGTGTTAGATTAGAGTAGATTAAACATTTATTGTCAGCAGTAGTAACTATCT 1851

QY      453 GlyTyrSerLeuAsnAsnThrLeuIleIlePheAspArgIleArgGluAsp---ArgGln 471
DB      1852 GGTATTATCAATCAATGATACAACTTGTGACCGTGTTCGAGAAATCTGCATAAA 1911

QY      472 AlaAsnLeuPheThr-----ProMetHisValLeuValAsnAspAlaLeuGlnLys 488
DB      1912 GTTAAAGTAAATGACATCTGATCAATTCATGATATAGTCAACCGCTCTATTAGACAA 1971

QY      489 ThrPheSerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMetLeuLeu 508
DB      1972 ACTATGACACGTTCTATTATATACAGTGTGACTGTAGTTGTTAGTTGCAATATTA 2031

QY      509 PheIleGlyGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyLeuLeu 528
DB      2032 ATATTAGGTGCACCAACAATTTAATTTCTTTAGCATTAATTTGATTTATTTATCT 2091

QY      529 GlyThrLeuSerSerLeuTyrIleAlaProProLeu 540
DB      2092 GGTGATTCTTCCTCAATTTTCATTTGCTGTACCAATTA 2127

RESULT 12
ABN91389
ID ABN91389 standard; DNA; 2301 BP.
XX
XX AC ABN91389;
XX
XX DT 24-JUL-2002 (first entry)
XX
XX DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:852.
XX
XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX
XX OS Staphylococcus epidermidis.
XX
XX PN US6380370-B1.
XX
XX PD 30-APR-2002.
XX
XX PF 13-AUG-1998; 98US-0134001.
XX
XX PR 14-AUG-1997; 97US-055779P.
XX
XX PA 08-NOV-1997; 97US-064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.

```

PI Doucette-Stamm LA, Bush D;  
 XX WPI; 2002-381255/41.  
 DR P-PSDB; ABP38844.  
 XX

Novel isolated nucleic acid encoding a *Staphylococcus epidermidis* polypeptide, useful for diagnosing and treating bacterial infections -  
 XX  
 XX Disclosure; SEQ ID 852; 267pp; English.

XX ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading frame (ORF) nucleic acid sequences which encode the amino acid sequences given in ABP35124 to ABP37960. The *S. epidermidis* sequences have antibacterial activity and can be used in gene therapy. The sequences can also be used in the diagnosis and treatment of bacterial infections, particularly *S. epidermidis* infections. The sequences can be used to screen for compounds able to interfere with the *S. epidermidis* life cycle or inhibit *S. epidermidis* infection.  
 CC N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the CC USPTO web site.  
 XX

SQ Sequence 2301 BP; 819 A; 333 C; 403 G; 746 T; 0 other;

Alignment Scores:  
 Pred. No.: 7,49e-62 Length: 2301  
 Score: 646.50 Matches: 164  
 Percent Similarity: 51.22% Conservative: 129  
 Best Local Similarity: 28.67% Mismatches: 200  
 Query Match: 23.27% Indels: 79  
 DB: 24 Gaps: 12

US-09-868-987-14 (1-552) x ABN91389 (1-2301)

Qy 2 ValSerSerProLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLysPhe 21  
 Db 625 GTAGACCAACCTATTAAATCTAGTAGTCTGTAATTTTCAGGTGGCTTCAATGGGAAAAA 684

Qy 22 ThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheVal 41  
 Db 685 GGTGTTGAAGAACGAAACAATAGCTGAGTTATTAAATGCCGCTCATTACCAAGTTGAT 744

Qy 42 ProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysLysGlnCysThrGln 61  
 Db 745 TTAAAGAAATTTACTCTAACTCTGTTGGTGCAACAATTTGCTCAAGATGCTCTGTGAAG 804

Qy 62 GlyIleLeuSerAlaCysCysGlyLeuAlaMetLeuLeuValLeuMetSerValTyr 81  
 Db 805 ACCATGTTGCATCAATTTGATGATAGTATAGCATTAATTTATTATTATGCTTGGTTCTAT 864

Qy 82 ArgPheGlyGlyValIleAlaSerGlyValValLeuLeuLeuAsnLeuLeuIleTrpAla 101  
 Db 865 CGTTTGGCTGTTTGTAGTTGCAATCATTTGCCCTTAACCACTTATATTTTAACTTTAGTC 924

Qy 102 AlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleValLeuAla 121  
 Db 925 GCATTCAATTTTCATCAGGTGTATTAACTCTACCTCGATGGATGGCGCATTTAGGT 984

Qy 122 MetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPheLeu 141  
 Db 985 GTAGGTATGGCTGTCGATGCAATATCATATGATGATGAACGTTATTAAGATGAAGA 1044

Qy 142 LeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAlaIle 161  
 Db 1045 ATTGGACGACGCTTAAACAGCGGTATTCAAAGCAAAATAAAGTTTCATTCTTAACTATA 1104

Qy 162 PheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuLeuPhePheLeuAspThrGly 181  
 Db 1105 TTGTATCCAACTTAAACAACCTGTCATCGTCGACGTGCTTCTTCTTTGGAGAAAGT 1164

Qy 182 ProLeuLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThra 201  
 Db 1165 TCAGTCAAGGCTTCGCAACCATGTTACTCTTAGTATTATTTAATGATATTGTGAACCGCA 1224

Qy 202 LeuPheMetThrLysPhePheMetLeu----- 211  
 Db 1225 GTATTTCTTCAAGAGGGTGTGTATCATTTACTGCTATCTTCAAACTCTTTTAAAAACAA 1284

Qy 212 ---TrpMetAsnLysThrGlnHisThrGlnLeuHisMetMetAsnLysPheValGlyIle 230  
 Db 1285 TACTGGTTATTTGGTGTAAAGAGATGGATAGACATGATATTAAATGAAGTAAAGATGTA 1344

Qy 231 LysHisAsp-----PheLeuArgGlyCysLysLysLeu 241  
 Db 1345 ---CATGATTTAAAAACATCATATGAAGGTTAAACTTGTTAATATTAGTAAAGCCACTT 1401

Qy 242 TrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrp 261  
 Db 1402 ATTTCTACTTAGTATTTAAATTTATTTGGTTGATATATCATTTCAATATTTAAATTA 1461

Qy 262 AsnSerValLeuGlyMetAspPheLysGlyGlyTyrAlaPheThrPheAsnProLysGlu 281  
 Db 1462 AAC-----TTAGTATTGATTTCTCATCGGAAACAGACAGATATTCAATCTAAAAAT 1515

Qy 282 HisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGluAla 301  
 Db 1516 -----GCTATAACACACAGACAGGTTGAGAAAACTGTAAAAATCAGTT 1557

Qy 302 GlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLysIle 321  
 Db 1558 GGATTTGGAACACAGAT-----CAATACAGATTAATGTGTAGTGGAAATAA----- 1602

Qy 322 TyrPheSerAspLysAlaLeuSerTyrThrLysGlnIleArgAlaSerLeuLysLeu 341  
 Db 1603 -----AATGCCACAGTTCACTTTAAAAAAGATTTATCACGT--- 1638

Qy 342 ThrIleMetSerTrpArgTyrCysGlyIleValValArgAsnArgProArgPheLeuTyr 361  
 Db 1639 -----GAG 1641

Qy 362 GlyAsnSerLysArgAsnAlaLysPheTrpSerLys----- 373

Db 1642 GAAGACAAATAAATAAGTCTAAGTGAATTTGGAATTTGGAGATTAATCCACAAATAAT 1701

Qy 374 ---ValSerSerLysLeuSerLysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGly 392  
 Db 1702 ACCGTTTCACTCTCATAGCCCAAGAGTAGTAAAAATGCTCTAACTGCATTAATACIT 1761

Qy 393 AlaLeuAlaIleLeuLeuLeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPheSer 412  
 Db 1762 GCTTCTATAGGCATTTATTTATCTATTTTCACTAAGATTTGAATGGCTATGGCTCTATCT 1821

Qy 413 AlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHis 432  
 Db 1822 TCTGTACTTGCATTTATCATGACGTATTT-----ATCATCATTTGCAATC 1866

Qy 433 PhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeu 452  
 Db 1867 TTTAGTTTGTATTAGATTAGATAGATTTTAACTTTATTTCAGCAGATTTAACTATCGTT 1926

Qy 453 GlyTyrSerLeuAsnAsnThrLeuIleLeuPheAspArgIleArgGluAsp---ArgGln 471  
 Db 1927 GGTATTATCAATCAATGATACAAATCGTAACTTTTCGACCCGTGTTCGAGAAAAATCTGCATAA 1986

Qy 472 AlaAsnLeuPheThr-----PromethHisValLeuValAsnAspAlaLeuGlnLys 488  
 Db 1987 GTTAAAGTAAATTACGCATATCTGATCAAAATTTGATGATATAGTCAACCGCTCTCTATTAGACAA 2046

Qy 489 ThrPheSerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMetLeuLeu 508  
 Db 2047 ACTATGACACAGTTCTATTAAATACAGTTGTGACTGTAGTTGTAGTTGCAATATTA 2106

Qy 509 PheIleGlyGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeu 528  
 Db 2107 ATATTAGGTGACCAACCAATATTTAAATTTCTCTTAGCATTTACTAATTTGGATTATTATCT 2166



QY 362 GlyAsnSerLysArgAsnAlaLysPheTrpSerLys----- 373  
 Db 1917 GAAGACAATAAATTAAGTGCTAAGGTGAATCTGAATTTGGAGATAATCCACAATAAT 1858  
 QY 374 ---ValSerSerLysLysLysMetArgTyrGlnAlaThrIleGlyLeuLeuGly 392  
 Db 1857 ACCGTTTCACCTCTCATAGGCAAGAGCTAGCTAAATGCTGTAACGTCAATTAATCTT 1798  
 QY 393 AlaLeuAlaIleLeuLeuTyrValSerLeuArgPheGluTrpGlnTyrAlaPheSer 412  
 Db 1797 GCTTCTATAGCAATTAATATCTATGTTCTCACTAAGATTGGAATGGGTATGGTCTATCT 1738  
 QY 413 AlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHis 432  
 Db 1737 TCTGTACTGTCATTAATACATGACGTATTT-----ATCATCATTTGCCAATC 1693  
 QY 433 PhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeu 452  
 Db 1692 TTTAGTTTGTAGATTAGATGATTTTACATTTATTGACGAGTATTAACTATCGTT 1633  
 QY 453 GlyTyrSerLeuAsnThrLeuIlePheAspArgIleArgGluAsp---ArgGln 471  
 Db 1632 GGTATTCAATCAATGATACAAATCGTAACTTTCCACCGTGTTCGAGAAATCTGCATAA 1573  
 QY 472 AlaAsnLeuPheThr-----PromethisValLeuValAsnAspAlaLeuGlnLys 488  
 Db 1572 GTTAAGTAATTAATCGCATCTACTGATCAAAATTCATGATATAGTCAACCGCTCTATTAGACAA 1513  
 QY 489 ThrPheSerArgThrValMetThrAlaThrThrLeuSerValLeuLeuMetLeuLeu 508  
 Db 1512 ACTATGACACGTTCTTATTAATACAGTGTGACTGTAGTTGTTAGTTGTCGAATTA 1453  
 QY 509 PheIleGlyGlySerValPheAsnPheAlaPheIleMetThrIleGlyIleLeuLeu 528  
 Db 1452 ATATTAGTGCCACCAACAATATTAATTTCTTTTAGCATTACTAATTTGGATTATATCT 1393  
 QY 529 GlyThrLeuSerSerLeuTyrIleAlaProLeu 540  
 Db 1392 GGTGTATTCGTCAATTTTTCATTGCTGTACCATT 1357  
 RESULT 14  
 AAH54228/C  
 ID AAH54228 standard; DNA; 3705 BP.  
 XX AAH54228;  
 AC  
 XX  
 DT 03-SEP-2001 (first entry)  
 XX  
 XX S. epidermidis genomic polynucleotide sequence SEQ ID NO:3592.  
 XX  
 KW Staphylococcus epidermidis SRI strain; infection; diagnosis;  
 KW vaccination; endocarditis; ds.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN WO200134809-A2.  
 XX  
 PD 17-MAY-2001.  
 XX  
 PF 09-NOV-2000; 2000WO-US30782.  
 XX  
 PR 09-NOV-1999; 99US-0164258.  
 XX (GLAX ) GLAXO GROUP LTD.  
 XX  
 XX Pi Kimmerly WJ;  
 XX  
 DR WPI; 2001-316495/33.  
 XX  
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
 XX useful for vaccinating against infections, e.g. endocarditis -  
 XX

PS Claim 8; Page 1183-1184; 2188pp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAG81454 to AAG8120, from Staphylococcus epidermidis.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH5090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH5091 to  
 CC AAH5098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO.4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.  
 XX

SQ Sequence 3705 BP; 1234 A; 621 C; 509 G; 1341 T; 0 other;

Alignment Scores:  
 Pred. No.: 1,57e-61 Length: 3705  
 Score: 646.50 Matches: 164  
 Percent Similarity: 51.22% Conservative: 129  
 Best Local Similarity: 28.67% Mismatches: 200  
 Query Match: 23.27% Indels: 79  
 DB: 22 Gaps: 12

US-09-868-987-14 (1-552) x AAH54228 (1-3705)

QY 2 ValSerSerProIleLeuAsnValProLeuLysAsnHisAlaSerValSerGlyLysPhe 21  
 Db 2211 GTAGACCAACCTATTAAATCTAGTAGTGTGAATTTCAAGTGGCTTCAATGGGAAAAA 2152  
 QY 22 ThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSerPheVal 41  
 Db 2151 GGTGTGAAGAGCGAAACAAATAGCTGAGTTAATATGCGCGCTCAATACCACTTGAT 2092  
 QY 42 ProGluValLeuSerGluThrIleSerSerAspLeuLysGlnCysThrGln 61  
 Db 2091 TTAAGAGAAATTTACTCTAATCTGTGTGTGCACAAATTTGGTCAAGATGCTCTTTGTAAG 2032  
 QY 62 GlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerValTyrTyr 81  
 Db 2031 ACCATGTTGTCATCAATTTAGTAGTAGCATTAATTTATTTATTTATTTGCTTCTAT 1972  
 QY 82 ArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIleTrpAla 101  
 Db 1971 CGTTTGCCTGGTTAGTTCGAATCATTCGCTTAAACCACTATATATTTATTTAAGTGC 1912  
 QY 102 AlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaClyIleValLeuAla 121  
 Db 1911 GCATTCATTTTCATATCAGGTGTATTAACTCTACCTGGATTTGGCGCATTTAGTTTAGGT 1852  
 QY 122 MetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGluPheLeu 141  
 Db 1851 GTAGGTATGGCTGTCGATCCCAATATCATATGATGATGATTAAGATTAAGACTAAGA 1792  
 QY 142 LeuSerGlnSerLeuLysSerValGluLysGlyTyrThrLysAlaPheGlyAlaIle 161  
 Db 1791 ATTGACGACGCTTAAACAAAGCTATTCAAAAGCAATATAAAAGTTTCATTTCTTAACATA 1732  
 QY 162 PheAspSerAsnLeuThrThrValLeuAlaSerAlaLeuPhePheLeuAspThrGly 181  
 Db 1731 TTGTATTCCAACTTAAACAACTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1672  
 QY 182 ProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPheThrAla 201  
 DB: 12 Gaps: 12



Db 1671 TCAGTCAAGGCTTCGCAACCATGTTACTCTTAGGTATTTAAATGATATTTGTAACCGCA 1612  
 Qy  
 Db 202 LeuPheMetThrLysPhePheMetLeu----- 211  
 Db 1611 GTAATTCCTTCAAGAGGGTGTATTCATTTACTGGTATCTTCAAACTCTTTAAAAAACA 1552  
 Qy 212 ---TTPMetAsnLysThrGlnHisThrGlnLeuHisMetMetAsnLysPheValGlyIle 230  
 Db 1551 TACTGTTATTGGTGTAAAGAGGATAGACATGATATTAATGAAGGTAAGATGTA 1492  
 Qy 231 LysHisAsp-----PheLeuArgGlyCysLysLysLeu 241  
 Db 1491 ---CATGATTTAAAAAACATCATATGAAAGGTTAAACTTTTGTAAATTAGCTAAGCCACTT 1435  
 Qy 242 TrpAlaValSerGlySerValPheLeuLeuGlyCysValAlaLeuGlyPheGlyAlaTrp 261  
 Db 1434 ATTTCACCTAGTATTTAATGTAATTAATGGTTGATTATCATTTCAATATTTAAATA 1375  
 Qy 262 AsnSerValLeuGlyMetAspPheLysGlyTyTyAlaPheThrPheAsnProLysGlu 281  
 Db 1374 AAC-----TAGGTATTGATTTCTCATCCGGAACAAGCAGATATTCATCTAATAAT 1321  
 Qy 282 HisGlyIleSerAspValAlaGlnMetArgGlyLysValValHisLysLeuGlnGluAla 301  
 Db 1320 -----GCTATAACAACAAGCAGGTTGAGAAAACTGTAATAATCAGTT 1279  
 Qy 302 GlyLeuSerSerArgAspPheArgIleGlnThrPheGlySerSerGluLysIleLysIle 321  
 Db 1278 GGATTTGGAACCCAGAT-----CAATACAGATTAATGGTAGTGGAAATAA----- 1234  
 Qy 322 TyrPheSerAspLysAlaLeuSerTyTyThrLysGlnIleArgAlaSerLeuLysLeu 341  
 Db 1233 -----AATCCACAGTTCAGTTTAAAAAGATTTATCAGT--- 1198  
 Qy 342 ThrIleMetSerTrpArgTyTyCysGlyIleValValArgAsnArgProArgPheLeuTy 361  
 Db 1197 -----CAG 1195  
 Qy 362 GlyAsnSerLysArgAsnAlaLysPheTrpSerLys----- 373  
 Db 1194 GAAGACAATAAATTAAGTGCTAAGTGAAATCTGAATTTGGAGATAAATCCAAATTAAT 1135  
 Qy 374 ---ValSerSerLysLysLysMetArgTyTyGlnAlaThrIleGlyLeuLeuGly 392  
 Db 1134 ACCGTTTCACCTCTCATGGCAAGCAGTAGCTAAATAATGCTGAACGTCAATATATCTT 1075  
 Qy 393 AlaLeuAlaIleLeuLeuTyTyValSerLeuArgPheGluTrpGlnTyAlaPheSer 412  
 Db 1074 GCTTCTATAGGCATTATATCTATCTGTTCACTAAGATTTGAATGCGGTCTATCT 1015  
 Qy 413 AlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeuPheIleAlaHis 432  
 Db 1014 TCTGTACTGTCATTATATCATGACGTATTT-----ATCATCTTGCATC 970  
 Qy 433 PhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMetThrValLeu 452  
 Db 969 TTAGTTGTTTGTAGTTAGATGATGATTAACATTTATTCGCGCAGTATTACATCTGTT 910  
 Qy 453 GlyTyTySerLeuAsnAsnThrLeuIlellePheAspArgIleArgGluAsp---ArgGln 471  
 Db 909 GGTATTCAATCAATGATACATCTGTAATTTTCGACCGGTTCGAGAAAACTGCATATA 850  
 Qy 472 AlaAsnLeuPheThr-----ProMetHisValLeuValAsnAspAlaLeuGlnLys 488  
 Db 849 GTTAAGTAATTAACCATCTGATCAAAATTTGATGATATAGTCAACCGCTCTATTAGACA 790  
 Qy 489 ThrPheSerArgTyTyValMetThrThrAlaThrLeuSerValLeuLeuMetLeuLeu 508  
 Db 789 ACTATGACACGGTCTTATTAATACAGCTGTGCTAGTTGTAGTTGTAGTTGCAATATA 730  
 Qy 509 PheIleGlyGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyLeuLeu 528  
 Db 729 ATATTAGGTGCACCAACAATAATTTAATTTCTTCTTAGCATTACTAATTTGGATTATATCT 670

Qy 529 GlyThrLeuSerSerLeuTyTyIleAlaProProLeu 540  
 Db 669 GGTGATTCTTCGTCATTTTTCATTTGCTGTACCAATTA 634

## RESULT 15

AAA81457/C

ID AAA81457 standard; DNA; 49646 BP.

XX AAA81457;

AC AAA81457;

XX 04-DEC-2000 (first entry)

XX N. meningitidis partial DNA sequence gnm\_5 SEQ ID NO:5.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;

XX antigen; vaccine; diagnosis; infection; antibacterial; identification;

XX Meningococcus B; MenB; ds.

XX Neisseria meningitidis.

XX WO200022430-A2.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-US23573.

XX 09-OCT-1998; 98US-0103794.

XX 30-APR-1999; 99US-0132068.

XX (CHIR ) CHIRON CORP.

XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC; Scarlato V;

XX Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Rappuoli R, Pizza M;

XX WPI; 2000-318079/27.

Isolated nucleotide sequences of *Neisseria meningitidis* which can be used in the diagnosis and treatment of *N. meningitidis* infection and other *Neisseria* infections, for example, *N. gonorrhoea* -

Claim 7; Page 274-288; 1760pp; English.

The present invention describes methods of obtaining immunogenic proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414 represent specifically claimed *Neisseria meningitidis* genomic DNA sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF sequences, which are all used in the exemplification of the present invention. The nucleic acid sequences, protein sequences, and antibodies against them, can be used in the manufacture of a composition. The composition can be used as a medicament (or in the manufacture of a medicament) for treating, preventing or diagnosing infection due to *Neisseria* bacteria. For example, some of the identified proteins could be components of vaccines against *Meningococcus B*; against all serotypes; and/or against all pathogenic *Neisseria*. Identification of sequences from the bacterium will also facilitate production of biological probes, particularly organism-specific probes. Attempts to make efficacious *Meningococcus B* vaccines have failed mainly due to antigen tolerance. Multivalent vaccines have also been tried but none have successfully overcome antigenic variability. The provision of further, complete sequences may provide an opportunity to identify secreted or surface exposed proteins that may be presumed targets for the immune system and which are not antigenically variable or at least more conserved than other more variable regions.

XX Sequence 49646 BP; 12331 A; 14486 C; 11862 G; 10966 T; 1 other;

Alignment Scores:

```
Pred. No.: 1.25e-56 Length: 49646
Score: 618.50 Matches: 160
Percent Similarity: 51.19% Conservative: 140
Best Local Similarity: 27.30% Mismatches: 209
Query Match: 22.26% Indels: 77
DB: 21 Gaps: 15

US-09-868-987-14 (1-552) x AAA81457 (1-49646)

QY 1 MetValSerSerProIleLeuAsnValProLeuLys---AsnHisAlaSerValSerGly 19
Db 31089 GTTGTAAACCGCGCGGTTATCCGTACTGCCATTACCGCGGACGCGTGGAATAATTCGGA 31030
QY 20 LysPheThrHisArgGluValSerLysLeuAlaSerAspLeuLysSerGlyAlaMetSer 39
Db 31029 AGCATGACGACGCGGAAGCAATGATACGTCCTTTGCTGTGCGCGGTCCTCTGCC 30970
QY 40 PheValProGluValLeuSerGluGluThrIleSerSerAspLeuGlyLysGlnCys 59
Db 30969 GCACCGATGCAGATTGTCGAAGAACGTAACCATCGTCGCTCTTTGGGTAAAGGAACATC 30910
QY 60 ThrGlnGlyIleIleSerAlaCysCysGlyLeuAlaMetLeuIleValLeuMetSerVal 79
Db 30909 GAAAAAGGCTTCCATTCCACTTTATGGGGTTTCCCATCGTGTGTCATTTCATCGTGGT 30850
QY 80 TyrTyrArgPheGlyGlyValIleAlaSerGlyAlaValLeuLeuAsnLeuLeuIle 99
Db 30849 TACTATCGTCGATGGGTTCTTTTCTACCATTCGATTGAGTGCAACATACACTGTTCTTA 30790
QY 100 TrpAlaLeuGlnTyrLeuAspAlaProLeuThrLeuSerGlyLeuAlaGlyIleVal 119
Db 30789 ATCGGTATTTTCTGCCATCAGCAACGTTGACGTTACCGGGTATGGCGCGCTGGCG 30730
QY 120 LeuAlaMetGlyMetAlaValAspAlaAsnValLeuValPheGluArgIleArgGluGlu 139
Db 30729 TTGACTTTGGGTATGGCAATCGATCCCAACGCTTGATTAAACGAACGATATCCGCGAAGA 30670
QY 140 PheLeuLeuSerGlnSerLeuLysLysSerValGluLysGlyTyrThrLysAlaPheGly 159
Db 30669 TTGGCGTCCGCGGTGCGCCGACGAGCATCAATCTCGGTTTCCAAACGCGATGGCG 30610
QY 160 AlaIlePheAspSerAsnLeuThrValLeuAlaSerAlaLeuLeuPheLeuAsp 179
Db 30609 ACCATTGTCGATTGCAACCTGACTTCGCTGATTGCGGTATCGCGCTTTTGGTATTCGGT 30550
QY 180 ThrGlyProIleLysGlyPheAlaLeuThrLeuIleLeuGlyIlePheSerSerMetPhe 199
Db 30549 TCCGCGCGGTACGCGGTTTTCGCGTGTACACTGTTTGGGTATTCGACTTCGATGTAT 30490
QY 200 ThrAlaLeuPheMetThrLysPhePheMetLeuTrpMetAsnLysThrGlnHisThr 219
Db 30489 TCATCCGTCGTCGATTCCGCGGCTGGTCAATCTGTGGTACGACGACGACGCAAAATG 30430
QY 220 GlnLeuHisMetMetAsnLysPheValGly----- 229
Db 30429 CAG-----AATATTTCATTGTTGCGGTGGAAGCCGAAAGCCGAAATGGCA 30382
QY 230 ----- 30381
Db 30381 GGAGGCAAGGAGTAAGCTATGGAACCTTTTAAATCAACGCGATATTCGTTTATGAGC 30322
QY 237 GlyCysLysLysLeuTrpAlaValSer-----GlySerValPheLeuLeu 251
Db 30321 TACCGCAACATGACGACCTTCATTTCGTTGATTACGTTTATCGTGGCGGTCTTTTGG 30262
QY 252 GlyCysValAlaLeuGlyPheGlyAlaTrpAsnSerValLeuGlyMetAspPheLysGly 271
Db 30261 GTTACCAGAGGTCGAATTC-----TCTGTGCAATTTACCGGC 30223
QY 272 GlyTyrAlaPheThrPheAsnProLysGluHisGlyIleSerAspValAlaGlnMetArg 291
Db 30222 GGT-----ACGGTAATGGAAGTCCAATATATCAGCAG 30193
```

```
QY 292 GlyLysValValHisLysLeuGlnGluAla-----GlyLeuSerSerArgAspPheArg 309
Db 30192 GGTGCGGATGTCATTAAGATGCGCGAAGCGCTCGATACGCTGAAATAGGTGATGTACAG 30133
QY 310 IleGlnThrPheGlySerSerGluLysIleLysIleTyrPheSerAspLysAlaLeuSer 329
Db 30132 GTTCAGGCATTGGGTACGAACAAACATCATCGCTCGCTCGCAACAAA----- 30082
QY 330 TyrThrLysGlnIleArgAlaSerLeuLysLeuThrIleMetSerTrpArgTyrCys 349
Db 30081 -----GAAGGTGTACTTCCGCACAGTGTCCAATCAGGTTATGGAT----- 30040
QY 350 GlyIleValValArgAsnArgProArgPheLeuTyrGlyAsnSerLysArgAsnAlaLys 369
Db 30039 ---TTGCTGAAAAAAGACAGTCCGACGTT-----ACCTTGCGCAAGTCGAA 29995
QY 370 PheTrpSerLysValSerSerLysLeuSerLysMetArgTyrGlnAlaThrIleGly 389
Db 29994 TTT-----ATCGCGCCGCAAGTCGGTAGGAATGGTAAGTAATGGATTGATGGCT 29944
QY 390 LeuGlyAlaLeuAlaIleLeuLeuTyrValSerLeuArgPheGluTyrGlnTyr 409
Db 29943 TTAGGTTTGTCTGTATCGGCATCATTTTACCTGTGATGCTTTTGAATGCGGTTTT 29884
QY 410 AlaPheSerAlaValCysAlaLeuIleHisAspLeuLeuAlaThrCysAlaValLeuPhe 429
Db 29883 GCGGTATCTGCCATTATCGCAATATGCAGCATC-----GTGATTATT 29839
QY 430 IleAlaHisPhePheLeuLysLysIleGlnIleAspLeuGlnAlaIleGlyAlaLeuMet 449
Db 29838 CTCGGTGTCTTCCCTTCTTCCAATGGGAATTTTCGCTGACCGCTCTTGGCGGTATCCTT 29779
QY 450 ThrValLeuGlyTyrSerLeuAsnAsnThrLeuIleIlePheAspArgIleArgGluAsp 469
Db 29778 GCGGTATGGGCTATTCTGTGAACGAATCCGTCGCTCTTCGACCGTATCCGTTGAAAC 29719
QY 470 -----ArgGlnAlaAsnLeuPheThrProMetHisValLeuValAsnAspAlaLeuGln 487
Db 29718 TTCCGCAAGCCGCGATGCGCGACATGCGTCCGCGAAGTCATCCACACGCGATTACC 29659
QY 488 LysThrPheSerArgThrValMetThrThrAlaThrThrLeuSerValLeuLeuMetLeu 507
Db 29658 GCAACGATGAGCGCGACCATCATTCACCGGTTCCAGCGGCGATGGTGTATCCATG 29599
QY 508 LeuPheIleGlyGlySerSerValPheAsnPheAlaPheIleMetThrIleGlyIleLeu 527
Db 29598 CTGGTGTTCGGCGGTGCGGCTTTGACGCGCTTTTCTATGGCGTTGACCATTTGCGATCGTG 29539
QY 528 LeuGlyThrLeuSerSerLeuTyrIleAlaProProLeuLeuLeu---PheMetValArg 546
Db 29538 TTCGGCATTTATTCTTCGCTATTGGTTGCGACCGCGCTTCTGCTAATGTTGCGTTTGAGC 29479
QY 547 LysGluAsnArgSerLys 552
Db 29478 CGCGACAATATCGGTAAA 29461
```

Search completed: October 28, 2003, 16:14:11

Job time : 2871 secs